

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run On: June 20, 2003, 02:47:58 ; Search time 6394 Seconds
(without alignments)
17232.284 Million cell updates/sec

Title: US-09-715-876-7
Perfect score: 3786
Sequence: 1 atgttcaacaattacatt.....tgtctgttcatttagtga 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl :
1: gb_ba :
2: gb_htg :
3: gb_in :
4: gb_om :
5: gb_ov :
6: gb_pat :
7: gb_ph :
8: gb_pl :
9: gb_pr :
10: gb_ro :
11: gb_sts :
12: gb_sy :
13: gb_un :
14: gb_vl :
15: em_ba :
16: em_fun :
17: em_hum :
18: em_in :
19: em_mu :
20: em_om :
21: em_or :
22: em_ov :
23: em_pat :
24: em_ph :
25: em_pl :
26: em_ro :
27: em_sts :
28: em_un :
29: em_vi :
30: em_htg_hum :
31: em_htg_inv :
32: em_htg_other :
33: em_htg_mus :
34: em_htg_pln :
35: em_htg_rnd :
36: em_htg_mam :
37: em_htg_vrt :
38: em_sv :
39: em_htgo_hum :
40: em_htgo_mus :
41: em_htgo_other :
Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	3786	100.0	3786	8	YSAALS1	L25902 Candida alb
2	1834.4	48.5	3360	8	CAU87956	U87956 Candida alb
3	1708.6	45.1	4383	8	AF051313	AF051313 Candida a
4	1526.4	40.3	4569	8	AF272027	AF272027 Candida a
5	1392.8	36.8	4723	8	AF025429	AF025429 Candida a
6	1188.8	31.4	3813	8	AF068866	AF068866 Candida a
7	1069.4	28.2	1071	6	AF044076	AF044076 Sequence
8	1069.4	28.2	1071	6	I65632	I65632 Sequence 1
9	1016.6	26.9	1404	8	CANALS2S1	AF024582 Candida a
10	1006	26.6	1404	8	CALALS1	AF024580 Candida a
11	905.8	23.9	4332	8	AF075293	AF075293 Candida a
12	874.4	23.1	6897	8	AF201684	AF201684 Candida a
13	842.8	22.3	1407	8	CANALS4S1	AF024584 Candida a
14	836.4	22.1	1407	8	CANALS42S1	AF024586 Candida a
15	805.4	21.3	1404	8	AF229989S1	AF229989 Candida a
16	731.8	19.3	1007	8	AF202529	AF202529 Candida d
17	507	13.4	995	8	AF202530	AF202530 Candida d
18	463.6	12.2	1013	8	AF201685	AF201685 Candida d
19	394	10.4	1299	8	AF075294	AF075294 Candida a
20	362.2	9.6	1058	8	AF201686	AF201686 Candida t
21	348.2	9.2	425	6	AR044080	AR044080 Sequence
22	348.2	9.2	425	6	I65636	I65636 Sequence 5
23	331.2	8.7	680	8	AF189016	AF189016 Candida a
24	202.4	5.3	1400	8	AB002099	AB002099 Candida t
25	198.6	5.2	13370	8	SPAPB2C8	AL590602 S.pombe c
26	170.8	4.5	172307	2	AC044842	AC044842 Homo sapi
27	141.6	3.7	377	8	AF211866	AF211866 Candida t
28	135.8	3.6	132449	9	AL365272	AL365272 Human DNA
29	134.4	3.5	20325	8	SPCPB16A4	AL591677 S.pombe c
30	132	3.5	380	8	AF211865	AF211865 Candida t
31	130.4	3.4	158615	2	AC117835	AC117835 Rattus no
32	126	3.3	185994	2	AC002042	AC002042 Homo sapi
33	125.8	3.3	175748	2	AC120669	AC120669 Rattus no
34	125.8	3.3	180903	2	AC125859	AC125859 Rattus no
35	121.6	3.2	21251	1	U23947	U23947 Mycoplasma
36	116	3.1	53352	5	AL592077	AL592077 Zebraphis
37	112.2	3.0	157	8	AF035757	AF035757 Candida a
38	111.8	3.0	180668	2	AC020857	AC020857 Mus muscu
39	111.6	2.9	666	8	AF413050S2	AF413051 Zea mays
40	111.2	2.9	22398	5	FRU271723	AJ271723 Fugu rubr
41	110.2	2.9	141017	2	AC116962	AC116962 Dictyoste
42	109.8	2.9	35412	8	SPCC188	AL049662 S.pombe c
43	109.8	2.9	38141	8	SPAC8A4	Z66569 S.pombe chr
44	109.6	2.9	127354	2	AC117014	AC117014 Rattus no
45	109.6	2.9	155019	2	AC117361	AC117361 Rattus no

ALIGNMENTS

RESULT 1
YSAALS1
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
YSAALS1
Candida albicans agglutinin-like sequence (ALS1) gene, complete cds.
L25902
L25902.1 GI:704426
Candida albicans.
Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
Hoyer, L.L., Scherer, S., Shatzman, A.R. and Livni, G.P.
Candida albicans ALS1: domains related to a Saccharomyces

[illegible]

Db 961 GGATCTAACGGTATTGTTATTGTTGGCTACTACCCAGAACAGTTACAGACAGTACTACCGCT 1020
Qy 1021 GTCACACTTACCATTCAAATCCCAAGTGTGATATAAACCAAAACAAATGAAATTTGCAA 1080
Db 1021 GTGACCACTTACCATTCGATCCTAACCGCGACAAACTAGACAAATTTGAATTTGAAA 1080
Qy 1081 CCTATTCCAAACCACTACCACCACTTCAATATGTTGGTGTGACTTTCCTATCTACT 1140
Db 1081 CCTATTCCAAACCACTACCACCACTTCAATATGTTGGTGTGACTTTCCTATCTACT 1140
Qy 1141 AAGACTGACCAATTTGGTGAACAGCTACTGTTATGTTGATGTCCTATCATATCACTACC 1200
Db 1141 AAACTGACCAATTTGGGGAACCTGCTACTGTTATGTTGATATTCCATATCACACTACC 1200
Qy 1201 ACAACTGTTTACCAAGTGAATGACAGGAACAATCACTACACACCAAACTCGTACCAATCCA 1260
Db 1201 ACTACTGTTTACCAAGTGAATGACAGGAACAATCACTTCCACCAACACACATCAATCCA 1260
Qy 1261 ACTGANTCAATGACACAGTGGTGTACAAGTTCACACTGCCAAATPCCAATGTTAGTACT 1320
Db 1261 ACTGACTCAATAGACACTGTCATTTGTACAAGTTCACACTGCCAAATPCCAATGTTACTACC 1320
Qy 1321 ACTGAATATTGGTCTCAGTCTCTTGTGTACAGACCACTACAGTTACTGCTCTCCAGGTGGT 1380
Db 1321 ACTGAATATTGGTCTCAGTCTCTTGTGTACAGACCACTACAGTTACTGCTCTCCAGGTGGT 1380
Qy 1381 ACCGATCTGTGATTATCAGAGAGCCACCAACCACTACTGTCACTACTACTGAATATTGG 1440
Db 1381 ACTGATCTGTTTAAATCAGAGACCCAGGACCACTACTGTCACTACTACCAACCGTACTGG 1440
Qy 1441 TCACAACTCTTTGCTACTACTACTGTTACTGCTCTCCAGGTGGTACTGACACTCAGTA 1500
Db 1441 TCAGAACTTACACTACTACTACTGTTACTGCTCTCCAGGTGGTACTGACACTCAGTT 1500
Qy 1501 ATTATCAGAGAACCAACCAATCCAACTGTCACTACCAACCGAGTATTGTTCAATCCTTT 1560
Db 1501 ATTATCAGAGAACCTTCCAAATCCAACTGTCACTACCAACCGAGTATTGTTCAATCCTTT 1560
Qy 1561 GCTACTACTACTACTACTACTGCTCTCCAGGTGGTACTGCTACTGATTAATATCAGAGAA 1620
Db 1561 ACTACCACTAGTACCTTCACTGCTCTCCAGGTGGTACTGATTAATATCAGAGAA 1620
Qy 1621 CCTCAAACCAACTGTCAACCACTACTGAATATTGTTGCTCCCACTTTACGCAACCAACT 1680
Db 1621 CCACCAACCAACTACTGTAAACCACTACTGAATATTGTTGCTCAACTCTTACACTACT 1680
Qy 1681 ACTGTGACTGCTCTCCAGGAGGCACTGACTGATTAATATCAGAGAACCAACCACT 1740
Db 1681 ACTGTGACTGCTCTCCAGGAGGCACTGACTGATTAATATCAGAGAACCAACCACT 1740
Qy 1741 ACTGTGACTGCTACTGATGTTGTCACAACTATGCTACCACTACTGATTAATATCAGAGAA 1800
Db 1741 ACTGTGACTGCTACTGATGTTGTCACAACTATGCTACCACTACTGATTAATATCAGAGAA 1800
Qy 1801 CCACAGGTGGTGTGACACTGTTATCATTTAGAGAGCCACCAACCACTGCTACTACT 1860
Db 1801 CCACAGGTGGTGTGATTCGGTTATCATTTAGAGAGCCCTCCAAATCCAACTGCTCAAC 1860
Qy 1861 ACTGAGTATGGTGTCAATGTTGCTACTACCACTGATTAATATCAGAGAACCAAGTGGC 1920
Db 1861 ACTGAGTATGGTGTCAATGTTGCTACTACCACTGATTAATATCAGAGAACCAAGTGGC 1920
Qy 1921 ACTGATCTGTTATCATTTAGGGAACCAACCACTACTGCTACCACTACTGATTAATATCAGAG 1980
Db 1921 ACCGATCTGCTTTATTTAGAGAACCAACCACTACTGTTAAACCACTACTGATTAATATCAGAG 1980
Qy 1981 TCTCAATCATATGCAACCACTACTACCAATTTACCGTCTCCAGGTGGTGAACCTGATCCGTT 2040
Db 1981 TCTCAATCTTATGCAACTACTACTACCAATTTACCGTCTCCAGGTGGTGAACCTGATCCGTT 2040
Qy 2041 CTTATCAGAGAGCCACCAACCACTACTGCTACTACTGATTAATATCAGAGAACCAATCATAT 2100
Db 2041 CTTATTAGGGAACCAACCACTACTGCTACTACTGATTAATATCAGAGAACCAATCATAT 2100

Qy 2101 GCTCAACCAACCACTGTTACTGCACCACTGTTGAAACCGATACACCGTCTTATTACAGAGAG 2160
Db 2101 GCTCAACCAACCACTGTTACTGCACCACTGTTGAAACCGATACACCGTCTTATTACAGAGAG 2160
Qy 2161 CCACAAACCACTACTGTCACCTACTACTGAATACTGGTGTCAATATATGCTACACACCACT 2220
Db 2161 CCACAAACCACTACTGTCACCTACTACTGAATACTGGTGTCAATATATGCTACACACCACTACT 2220
Qy 2221 ACTGTTACTGTCACCACTGTTGTCACCTACTACTGTTATCATTTAGAGAGCCACCAATCCA 2280
Db 2221 ACCATTACCGTCCACCTGGTGAGACCGATACCGTCTTATTAGGGAACCAACCAATCACT 2280
Qy 2281 ACAGTTACTACTACTGAATATTGGTGTCAACATCATTTGCCAACCAACCACTACTACTGCT 2340
Db 2281 ACAGTTACTACTACTGAATATTGGTGTCAACATCATTTGCCAACCAACCACTACTACTGCT 2340
Qy 2341 CCTCAGGTGGTGTGACACCTGTGATTTATGTA---AAGCATGTCAAGTTCAAGATT 2397
Db 2341 CCACCAAGGTGAACTGATCTGTTTAAATCAGAGAGCCACCAACCACTGTTTACCACCT 2400
Qy 2398 TCTACATCTCCTCAATCATATACCACTGATCATTTCCCATCTTTCCCGTCTCTATTATGTC 2457
Db 2401 ACCGAATACTGCTCTCAATCTTATACCACTGCTACTACCGTGTGACACCACTGCTGGA 2460
Qy 2458 AACAGCACCAACCTCCGATTTGTCACATTTGAATCTTCAATCATGAAATAC---TCCCTACT 2514
Db 2461 ACTGATCTGCTGATTTATTTATGACACCACTGTCAAGTTCAGAAATTTCTTCAATTTCTGCT 2520
Qy 2515 TCTATCAGTGTGATGGTGTGTTGCTTCTACAACCTTTGGTGTACTGAATCAGAGAACT 2574
Db 2521 CCTCATACCAACCACTACTGTTGTTGCTTACAACCTGTTGGTATTGAAACCAACCACT 2580
Qy 2575 ACTCAGAACTGATTTGCGAGTGTGTAAGAGTGTCTTAGATTTGTCAGTCTTCTGCTGCT 2634
Db 2581 ATTACAGAACTAGTGTGTAAGGTGATAAAGTGTGTTCTGTTGCTGTTCTTCTACTGCT 2640
Qy 2635 ATTGTCAACAAATCCAGATGACAAATGAACTCCTCAATCGTAACCTAGTACTGTTCTCTACTGCA 2694
Db 2641 ATTGTCAACAAATCCTTAATTAATATCGAACTCTCTATGTTTACTTAATCTGTTACTACTA 2700
Qy 2695 AGTCAATGCTGTGATCTACTTTTCAACTGATGTTATGTTGCTGATGCTGCTGCTGCTGCT 2754
Db 2701 ACCCAGAACTCCACTTCAACTCCCACTC---TGGTATTGTTTTCAGAGTCTGAGGATCT 2757
Qy 2755 GTTTCAAATCAGGAGTATCAGTTACACCGAACTTCTGTTTACAACTATTCAAACCTACT 2814
Db 2758 GTTGAACAGAACTCTTCTACTGTTTACTGCTCAACCAAAATCCCAAGTGTTCACCACT 2817
Qy 2815 CCAAAACCACTTATCTCTTCACTGATCATTTGACTGCTGTTCTTCAATTTCAAGTGT 2874
Db 2818 GAAAGTGAGGTTGATTTTACTACTAAAGGAACCAAGAAATGGTCTTATGNAATCACC 2877
Qy 2875 TCAGAAAGTGAAAGTAAAGTTTACATTTTACAGCAATGGAGACCAACAAAGTGGTACTCAT 2934
Db 2878 TCTACTAATGTCAAACTCAAGTATGGATGAAACTCTGGAATTTA---CTACTTCCACAGCT 2934
Qy 2935 GATTCAACATCTACTTCCACTGAAATTTGAA---ATTGTAAACCACTGTTCTACTAAAGTT 2991
Db 2935 GCTTCCACTTCTACTGATATTGAAATGAAACCACTAGCAACCAACCGTTTCCGTTGGAAGT 2994
Qy 2992 TTAACCACTGCTGTTTCTTCTTAAATCTGATTTGACTAGTGAACCAACCAATACCAAGAA 3051
Db 2995 TCATCGCTATCTTCTTCTAGTGTGATGAACTACTACTGTTTACTACTACTGCTGAA 3054
Qy 3052 CAACCACTTACTTACTACTTCAAACTCCACTCACTGAAAGATATACCACTACTCA 3111
Db 3055 TCAACCACTGCTTATGAAACCAACCACTAAATATGTTGTTGTTGTTGTTGTTGTTGTTGTT 3114
Qy 3112 CTTACAGGTGATTAAGGAGACAACTACTTCAACCACTCACTGTTTCCAGTCTGTTGCAACA 3171
Db 3115 ACTTCACTCCACTCAACCTCAACTGCTAATATGACTGTTTATTACTGTTGTTGTTGTTGTT 3174

```
QY 3172 ACTACTTTAGCATCTGCAAGTGAAGAAGACACAAAAGGGTTCTCATGCAATCAGCAT 3229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3175 TCACCAACCAATCTCAATCTCAATATATTTCTGATACCAACCAACTACAT 3232

RESULT 3
AF051313
LOCUS
DEFINITION
    AF051313
    Candida albicans agglutinin-like cell surface protein (ALS8) gene,
    complete cds.
    AF051313.2 GI:9625353
VERSION
    AF051313
KEYWORDS
    Candida albicans.
SOURCE
    Candida albicans.
    Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
    Saccharomycetales; mitosporic Saccharomycetales; Candida.
    1 (bases 1 to 4383)
    Leng, P., Lee, P.R., Wishart, J.A., Wu, H. and Brown, A.J.P.
    Sequence of the hypha-specific, agglutinin-like cell surface
    protein, ALS8 from Candida albicans
    Unpublished
    2 (bases 1 to 4383)
    Leng, P., Lee, P.R., Wishart, J.A., Wu, H. and Brown, A.J.P.
    Direct Submission
    Submitted (29-JUL-1999) Molecular and Cell Biology, University of
    Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen
    AB25 2ZD, UK
    On Aug 1, 2000 this sequence version replaced gi:4105850.
FEATURES
    source
        1..4383
        /organism="Candida albicans"
        /strain="ATCC10261"
        /db_xref="AFCC:10261"
        /db_xref="taxon:5476"
        566..>3840
        /gene="ALS8"
        566..>3840
        /gene="ALS8"
        /product="agglutinin-like cell surface protein"
        /note="minor form"
        574..579
        /gene="ALS8"
        /note="E-box"
        /bound_moiety="Efg1p"
        599..503
        /gene="ALS8"
        612..>3840
        /gene="ALS8"
        /product="agglutinin-like cell surface protein"
        /note="minor form"
        620..>3840
        /gene="ALS8"
        /product="agglutinin-like cell surface protein"
        /note="major form"
        630..>3840
        /gene="ALS8"
        /product="agglutinin-like cell surface protein"
        /note="minor form"
        697..3840
        /gene="ALS8"
        /function="involved in cell-cell adhesion"
        /note="hypha-specific cell surface glycoprotein; similar
        to Candida albicans strain 1161 agglutinin-like protein
        encoded by ALS3 gene sequence presented in GenBank
        Accession Number U87956; contains two fewer internal
        repeated units than the Candida albicans strain 1161
        gene"
        /codon_start=1
        /transl_table=12
        /product="agglutinin-like cell surface protein"
        /protein_id="AAD02580.1"
        /db_xref="GI:4105851"
```

```
/translation="MLQOYTLILLIYLSVATKTIITGVENSFNSLTWSNAATYHYKGP
TPTNVAVLGSIDGTASPGDTFLNMPCKVFKTTSTQVLDLTAHGKVAATCOFOAGE
EFMFSTLCTVSNLTSPSIRKALGTVTLPAFNVGSTGSSVLDLDESKFTAGTNTYTF
NDGKISINVDERSNPKGIDSLRNVIFSLNKVSTLFLVAPQCANGTSTGTMGFAN
DAYISQIDCSNIHVGIITKGLDNWNPVSEFSYTKTKSSNGIFITYKNVPAGYRPFV
TDSYATDVNSYTLISYANEYTCAGYWRAPFTLRWTGYNRSDAGSNGIIVATRTV
VDLPVHTTTTVPDNRDKTKTIEILKPIPTTITTSYGVPTSTVKAPIGETATVI
TTITGPGNTDVLIREPPNHTVTTEYWSYTTTSTFTAPGCTGTSVLIKEPPNP
TVTTEYWSYTTTSTAPGGTDSVLIKEPPNPTVTTEYWSYTTTSTFTAPGCTGTSV
GGTDLVLRPPNHTVTTEYWSYTTTSTFTAPGGTDSVLIKEPPNHTVTTEYWSY
SQSVATTTTITAPGETDVLIREPPNHTVTTEYWSYATTTITAPGGTDTVLI
REPNNHTVTTEYWSYATTTITAPGGTDTVLIKEPPNHTVTTEYWSYATTTIT
TVTAPGGTDTVLIYDTMSSEISFSRPHVNTHTLWSTTWIETKTIITETSCGDK
GCSWSVSYTRIVTIPNNLETVMVNTVDSITTESISQSPSGIFSESSVSVESTVT
TAQNPSPVTTESEVEFTKNGNPGYSPSHVKSMDENSESTETASTASTDIE
NATATGTSVEASSPIISSADETTITTTAESTSVIEQPTNNNGGKAPATSPST
TTTANNDSVITGTTSTNQSOSQSSQSSNSTQDTTLLSQQMTSSLVLSLHMLTTFD
GSGSVI
QHSWLCGLITLLSLFI"
BASE COUNT 1359 a 922 c 705 g 1397 t
ORIGIN
Query Match 45.1%; Score 1708.6; DB 8; Length 4383;
Best Local Similarity 84.6%; Pred. No. 1.1e-298;
Matches 1918; Conservative 0; Mismatches 349; Indels 0; Gaps 0;
QY 1 ATGCTTCAACAAATTTACATTTGTTATTCCTATATTTGTCAAATGCAAGTGCACAAATC 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 697 ATGCTACAAATATACATTTGTTACTCATATATTTGTCGGTGGCAAGTGCACAAATC 756
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 ACTGCTGTTTGTAGTATTTAAATTCATTAATGCTGCTCAATGCTGCTAAATTTATGCTTC 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 ACTGCTGTTTCAACAGTTTAAATTCATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCT 816
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 AAAGGCCAGGATACCCAACTTGGAATGCTGTTTGGGTGGTCTCTAGAGGTACCAAGT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 817 AAGGACCAGGAACCCCAACTTGGAATGCTGTTTGGGTGGTCTCTAGAGGTGCTACTAGT 876
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 GCCATCCAGGGGATACATTCACATTTGAATATGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 877 GCAAGTCGGGGAGATACATTCACATTTGAATATGCTGCTGCTGCTGCTGCTGCTGCTGCT 936
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 CAACATCTGTTGATTTAACTCCGATGCTGTTAAATATGCTGCTGCTGCTGCTGCTGCTGCT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 937 CAACATCTGTTGATTTGACTGCTCATGGTGAATATGCTGCTGCTGCTGCTGCTGCTGCTGCT 996
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 GGTGAAGAAATTCACAACTTTTCTACATTAACATGCTGCTGGAACGAGCTTTGAAATCA 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 997 GGTGAAGAAATTTATGACCTTTTCTACATTAACATGCTGCTGGAACAACTTTGACTTCCA 1056
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 361 TCCATTAAGGCATTTGGTACAGTTTACTTTACCAATTCATTCATCAATTTGGTGGACAGCT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1057 TCTATTAAGGCTTTGGGTACTGTCACCTTACCACCTTGCAATTCATTAAGTGGAACTGGT 1116
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 421 TCATCAACTGATTTGGAAGATTTCTAAATGTTTACTGCTGGTGGTACCACATACAGTTCATTT 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1117 TCTTCTGTTGATTTGGAAGATTTCTAAATGTTTACTGCTGGTGGTACCACATACAGTTCATTT 1176
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 481 AATGATGGTGATAAAGATATCTCAATGATGTTGAGTTTGAAGTCAACCGCTTGATGCCA 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1177 AATGATGGTGGAAGAAATCTCTATTAATGTTGATTTGAAGAGGTCAAAATTCGATCCA 1236
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 541 AGTGATATTTGATGCTTCAGAGTTTATGCAAGTCTCAATTAAGGTCAACTCTTTT 600
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1237 AAAGGGTACTTAAGTATCCAGAGTTTATACCAAGTCTCAACAAAGGTTCACAACTCTTTT 1296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 601 GTGGCACCACAAATGTGAATAAGTGTACACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1297 GTTGACCAACATGTGCAAAATGGTTACACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1356
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 661 GGTGACGTTGCTGATTTGCTGCTCAATATTCATATTTGTTGATCATCAAAAGGATTAATGAT 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1357 GGTGATGTTCAAAATTTGACTGTTTCAAAATATTCATGTTGGTATTTACAAAAGGATTAATGAT 1416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Qy	721	TGGAATATACCGGTTTCATCTGAATCATTTAGTTCACATAAAACTTGTACATCTAATGGA	780
Db	1417	TGGAATATACCGGTTTCATCTGAATCATTTAGTTCACATAAAACTTGTTCATCTAATGGT	1476
Qy	781	ATTCCAGATTAAATATCAAAATGACCTGCTGGTTATTCGTCATTTATTTGATGCTTATATT	840
Db	1477	ATCTTTATCACATATAAAAGTTCTTGC CGGTTATTCGTCCTATTTGTTGACGCTTATATT	1536
Qy	841	TC TGCTACAGATTTAAACCAATATACTTTAGCATATACCAATGATTATATCTTGTGCTGCC	900
Db	1537	TC TGCTACAGATTTAACTCGTACACCTTTGCTGATGCTTAATGAATATATCTTGTGCTGGT	1596
Qy	901	AGTCGCTCGAAAGTAAACCTTTTCACCTTTAAGATGAGCTGGATACAGAATAGTGATGCC	960
Db	1597	GGTTATGGCAACGTCACCTTTTCATATTAAAGATGGACTGGATACAGAAATAGTGATGCT	1656
Qy	961	GGATCAACCGTATTGTCATTGTTGGTTACAACTAGAACAGTTACAGACAGTACACACTGCT	1020
Db	1657	GGATCAACCGTATTGTTATTTGGTGCTACTACCAGAACAGTTACAGACAGTACTACGCC	1716
Qy	1021	GTCACTACTTTACCAATCCAAGTGTGTGATAAAACCAAAACAATCGAAATTTTGCAA	1080
Db	1717	GTCAACACCTTACCATTTCGATCTAAACCGCGACAAAACTAAGACAAATTGAAATTTTGAAA	1776
Qy	1081	CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGACTACTTCCATCTGACT	1140
Db	1777	CCTATTCCAACCACTACAATCACAAATCATATGTTGGTGACTACTTCCCTACCTGACC	1836
Qy	1141	AAGACTGCACCAATTTGGTGAACAGCTACTGTTATTGTTGATGTGCCATATCATCTACC	1200
Db	1837	AAAACGTCCACCAATTTGGGAAACTGCTACTGTTATTGTTGATATTCATATCACACTACC	1896
Qy	1201	ACAACGTGTTACCAGTGAATGGACGAGAACAAATCACTACCACCAACTCGTACCAATCCA	1260
Db	1897	ACTACTGTTACCAGTAAATGGACGAGAACAAATTTACTTCCACCAACACATACTAATCCA	1956
Qy	1261	ACTGATTCAATTCACACAGTGGTGACAAAGTTCCACTGCCAATCCAAGTGTAGTACT	1320
Db	1957	ACTGACTCAATGACACTGTTCATGTGACAAAGTTCCACTGGCCAAACCACTGTTACTACC	2016
Qy	1321	ACTGAATATTGGTCTCAGTCCCTTTGGTTACAACCACTACAGTTACTGCTCCTCCAGGTGGT	1380
Db	2017	ACTGAATATTGGTCTCAATCATTTGTTACCAACCAACCCACTTACTGGACCAACGAGAAC	2076
Qy	1381	ACCGATACTGTGATTTATCAGAGAGCCACCAACCACTACTGTCTACTACTAGTAATATGG	1440
Db	2077	ACTGATACTGTTTAAATCAGAGAACCAACCAACCACTACTGTCACTACAAACCGAGTATGG	2136
Qy	1441	TCACAATCCTTTGCTACTACTACTACTGTTTACTGCTCTCCAGGTGGTACTGACTCAGTA	1500
Db	2137	TCAGATCTTTACACTACCCTACTACTTTTCACCTGCTCTCCAGGTGAACTGACTCGGTC	2196
Qy	1501	ATTATCAGAGAACCAACCAATCCAACCTGTCACCTACAACCGAGTATTTGGTCTCAATCCTTT	1560
Db	2197	ATCATCAAGGAACCTCCAATTCACCTGTTTACAACTAGCGAGTACTGGTCAGAATCTTAC	2256
Qy	1561	GCTACTACTACTACAGTTACTGCTCTCCAGGTGGTACTGCACTCAGTAAATATCAGAGAA	1620
Db	2257	ACTACCACTAGTACCTTTCATCTGCTCTCCAGGTGGAACCTGACTCGGTCAATCATCAGGAA	2316
Qy	1621	CCTCCAAACCAACTGTCACCAACCACTGAAATTTGGTCCCAATCTTACGCAACCAACAAC	1680
Db	2317	CCACCAATCCAACTGTCACCTACACAGAGTACTGGTCACAACTCTTACACTACCACTACT	2376
Qy	1681	ACTGTGACTGCTCCTCCAGGAGCACTGACTCAGTAATTTATCAGAGAACCAACCAACCAAC	1740
Db	2377	ACTGTCAACGCTCCACCAAGGAGTACTGATACTGTCTTGTAGTCAGAGGCCACCAACCAAT	2436
Qy	1741	ACTGTCACTACTACTGAATACTGGTCACAATCATATATGCCCACCACTACCACTGTAATGCA	1800
Db	2437	ACTGTTACAATACCAGTACTGGTCACAATCTTACACTACAACCAACCACTGTAATGTC	2496

Qy	1801	CCACGAGTGGTACTGACACTGTTATCATTTAGAGAGCCACCAACACACTGTCACTACT	1861
Db	2497	CCACGAGTGGTACTGACACTGTTATCATTTAGAGAACCCACCAACACACTGTCACTACT	2556
Qy	1861	ACTGAGTATTGGTCTCAATCGTTTGGTCTACTACACAACTCTAACTGGTCCACCAAGTGGC	1920
Db	2557	ACTGAGTATTGGTCTCAATCTTACGCAACCACTACTACCAATTACCGCTCCACCTGGTGAG	2616
Qy	1921	ACTGATACTGTTATCATTTAGGGAACCAACCAACCCAACTCTGCACCACTCAATACTGG	1980
Db	2617	ACCGATACTGCTCTTATTAGAGAACCAACCAACCACTACTGTACCACTGAGTATTGG	2676
Qy	1981	TCTCAATCATATGCAACCACTACTACCATTTACCGTCCCACTGGTGAACACTGATACGGTT	2040
Db	2677	TCTCAATCCCTATGCACTACTACTACAATCACTGCTCTCCAGGTGAAACCGATACCGTT	2736
Qy	2041	CTTATCAGAGAGCCACCAACCACTACTGTCTCACTACTCAATCACTGGTCTCAATCATAT	2100
Db	2737	CTTATTAGGGAACCAACCAATCACTGTCTCACTACTCAATCACTGGTCTCAATCATAT	2796
Qy	2101	GCTACCAACCACTGTTTACTGCAACCACTGGTGAACCACTATACCGTCTCTTATCAGAGAG	2160
Db	2797	GCTACCACTACCACTGTAATCTGCACCACTGGTGGTACTGCACACTGTTCTTATCAGAGAG	2856
Qy	2161	CCACCAACCACTACTGTCACTACTACTGTAATCACTGGTCTCAATCATATGCTACAACACC	2220
Db	2857	CCACCAACCACTGTCACTACTACTGTAATCACTGGTCTCAATCATATGCTACAACACC	2916
Qy	2221	ACTGTTACTGCACCACTGGTGTACCGATACTGTTATCATTAGAGA	2267
Db	2917	ACTGTTACTGCACCACTGGTGTACCGATACTGTTATCATTAGAGA	2963
RESULT 4			
AF272027			
LOCUS	AF272027	4569 bp	mRNA linear PLN 24-OCT-2000
DEFINITION	Candida albicans agglutinin-like protein mRNA, partial cds.		
ACCESSION	AF272027		
VERSION	AF272027.1	GI:10952735	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
CDS			

Query Match	40.3%	Score 1526.4	DB 8	Length 4569
Best Local Similarity	77.7%	Pred. No. 8.8e-266		
Matches 1845	Conservative	0	Mismatches 531	Indels 0
			Gaps	0
QY	1	ATGCTTCAACAATTTACATTTGTTATTCCCTATATTTTGTCAATTCGAAGTCGAAGACCAATC	60	
DB	1	ATGCTTTTACAAATTTTGTGTAAGCCTCTGTGTATCAGTTGCTTACGGCAAAAGGTATT	60	
QY	61	ACTGGTGTGTTTGGATAGTTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTTAGCTTTC	120	
DB	61	ACGGGCATTTTCGATAGTTTAAATTCGTAACTTGGACCAATGCTGCTTCTTATTATAT	120	
QY	121	AAAGGGCCAGGATACCAACTTGGATGCTGTTTTGGTGGTTCCTAGATGTCACAGT	180	
DB	121	AGAGTCCAGCTAATCCTACTTGGACGGCTGTAAATAGGATGGTCTTTAGTAGAGCTACT	180	
QY	181	GCCAAATCCAGGGATACATTTACATTTGAATATGCCATGTGTGTTAAATATATACTACTCA	240	
DB	181	GCTAGTGGTGACACATTCAGCTTAGACATGCCATGTGTTTCAAAATTTATTACTGAT	240	
QY	241	CAAAACATCTGTTGATTTAACTGCCGATGGTGTAAATATGCTACTTGTCAATTTTATCTT	300	
DB	241	CAAAACGTCATTCATTTAGTTGCTGATGGTGCCTACTTATGCTACTTGTAAATTTGAACCT	300	
QY	301	GGTGAAGATTCACAACTTTTCTACATTAACATGACTGTGGAACGCGTTTGAATCA	360	
DB	301	GCCGAAGAGTTTACTACTTTTTCTAGTGTGTCACTGACTGTGACTACTCAAAATGACTGCT	360	
QY	361	TCCATTAAAGSCATTTGGTACAGTTACTTTACCAATTCGATTCAAATGCTGGTGGCAACAGT	420	
DB	361	GACCCAAAGCCATAGGAATGTAAACATTACCTTTCTCATTCAGTGTGGGGGATCAGGT	420	
QY	421	TCATCAACTGATTTGGAAGATTTCTAAATGTTTACTGCTGGTACCAATACAGTCACATTT	480	
DB	421	TCAGATGTTGATTTGGCAAAATCTCAATGTTTACTGCAGGAATCAATACAGTTACTTTT	480	
QY	481	AATGATGGTGAATAAGATATCTCAATGTAGTTGGTGAAGAAGTCAACCGTTGATPCCA	540	
DB	481	AATGATGGTGACACTAGCACTTTCCGCAACAGTTTGATTTTGAAAAATCAACCGTGGCTCC	540	
QY	541	AGTGCAATATTGATGCTTCCAGAGTTATGCCAAGTCTCAATAAAGTCACAACTCTTTT	600	
DB	541	AGCGATCGTATCTGTTGTCAAGAAATTTTACCAGTCTTTTCAACAGCATGAAGTCCTTTT	600	
QY	601	GTGGCACCACAATGTGAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAAC	660	
DB	601	CTTCCCCAAGAATGTGCAAAATGTTACTTCTGGTACAATGGATTTTCAACTGCTGGT	660	
QY	661	GGTGACGTTGCTATTGATTGCTCAAAATATTCATATTGGTATTCACAAAGAGTTAAATGAT	720	
DB	661	ACTGGTGCTACTATAGATTGTCCACAGTTTCATGTCGGGATATCAATGGGTTGAATGAT	720	

[illegible]

```

RESULT 5
AF025429
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
gene
CODS

```

[illegible]

QY 541 AGTGCATATTTGTATGCTTCCAGATTATGCGAAGTCTCAATAAGGTCAACAACCTCTTTT 600
DB 867 AGTGGTATTTGACTACTTCCAGATTATGCGAGTCTCAATAAAATGCTACTCTTTAT 926
QY 601 GTGGCACCACATGTGAATGGTTACACATCTGTACAATGGGGTTCTCCAGTAGTAAC 660
DB 927 GTGGCACCACATGTGAATGGTTACACATCTGTACAATGGGGTTCTCCAGTAGTTAT 986
QY 661 GGTGACGTGTGCTGATTTGCTCAATATTCATATTTGTTATCCACAAAGGATTAATGAT 720
DB 987 GGGGATGTTGCTGATTTGCTGATTTGCTGATTTGCTGATTTGCTGATTTGCTGATTT 1046
QY 721 TGAATATTCGCGTTTCTATCTGAATCATTTAGTTACACTAAACCTTTGATCTTAATGGA 780
DB 1047 TGAATATTCGCGTTTCTATCTGAATCATTTAGTTACACTAAACCTTTGATCTTTTGGT 1106
QY 781 ATTCAGATTAATATCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 1107 ATCTCTATACATATCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1166
QY 841 TCTGCTACAGATGTTTAAACCAATATCTTTAGCATATACCAATGATTTACTTGTCTGCTG 900
DB 1167 TCTCCTCAGATAAACCAGTATCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1226
QY 901 AGTCTGCTGCAAGTAAACCTTTCACTTTAAGATGGAGTGGATACAAGAATAGTGATGCC 960
DB 1227 GATTATGCGAGCATGCACTTTCACTTTAAATGGAGTGGATATAAGAATAGTGATGCC 1286
QY 961 GGATCTAAGGTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 1287 GGATCTAAGGTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1346
QY 1021 GTCAGTCTTTACCAATCAATCCAAAGTGTGATATAAACCACCAATCGAATTTTGCA 1080
DB 1347 GTCAGTCTTTACCAATCAATCCAAAGTGTGATATAAACCACCAATCGAATTTTGCA 1406
QY 1081 CCTATTCCAACCACTACCACACTTCTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
DB 1407 CCTATTCCAACCACTACCACACTTCTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1466
QY 1141 AGAGTGCACCAATTTGGTGAACAGCTACTGTTATTTGATGCTGCTGCTGCTGCTGCT 1200
DB 1467 AGAGTGCACCAATTTGGTGAACAGCTACTGTTATTTGATGCTGCTGCTGCTGCTGCT 1526
QY 1201 ACAACTGTTACAGTGAATGGACAGCAACATCACTACCAACCAACCACTCGTACCACCA 1260
DB 1527 ACAACTGTTACAGTGAATGGACAGCAACATCACTACCAACCAACCACTCGTACCACCA 1586
QY 1261 ACTGATTCATTTGACACAGTGGTGTGATAGTTCCACTGCGCAATCCAACTGTTAGTACT 1320
DB 1587 ACTGATTCATTTGACACAGTGGTGTGATAGTTCCACTGCGCAATCCAACTGTTAGTACT 1646
QY 1321 ACTGAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1647 ACCGAGTTTGGTCAAGTCAATTTACTAGTACTACCAATCAACCACTCTTTAAAGGT 1706
QY 1381 ACCGAGTGTGATTTACAGAGCCACCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1707 ACAGACTCAGTCAATTTAGAGAACCAACCAACCACTGCTGCTGCTGCTGCTGCTGCTGCT 1766
QY 1441 TCAGAACTCCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB 1767 TCAGAACTCCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1826
QY 1501 ATTATCAGAGAACCAACCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
DB 1827 ATTGTTAGAGAACCAACCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1886
QY 1561 GCTACTACTTACAGTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
DB 1887 GCTACTACTTACAGTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1946
QY 1621 CCTCCAAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680

DB 1947 CCACAAACCCCTACTGTGACTTACCACCAAGTTTGGTCAGAGTCTGATCTACTCTAG 2006
QY 1681 ACTGTGACTGCTCTCCAGGAGGCACTGACTCAGTAATATATCAGAGAACCAACCAAC 1740
DB 2007 ACCATCCACCAACAGCCAGAGGCACTGACTCAGTAATATATCAGAGAACCAACCAAC 2066
QY 1741 ACTGTGACTGCTGCTCTCCAGGAGGCACTGACTCAGTAATATATCAGAGAACCAACCAAC 1800
DB 2067 ACTGTGACTGCTCTCCAGGAGGCACTGACTCAGTAATATATCAGAGAACCAACCAAC 2126
QY 1801 CCACAAACCCCTACTGTGACTTACCACCAAGTTTGGTCAGAGTCTGATCTACTGAGACCAATCACT 1860
DB 2127 GGTCCAGAGGAAGAACAGACTCAGTCAATGTTAGAGAACCAACCAACCACTACTGTGACTACC 2186
QY 1861 ACTGAGTATTTGGTCTCAATCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
DB 2187 ACCGAGTTTGGTCTGAGTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2246
QY 1921 ACTGATCTGTTATCATTAGGGAACCAACCAACCACTGTCACCACTACTG 1972
DB 2247 ACTGATCTGTTATCATTAGGGAACCAACCAACCACTGTCCTTCTACTG 2298

RESULT 6
AF068866
LOCUS 3813 bp DNA linear PLN 29-JAN-2001
DEFINITION Candida albicans agglutinin-like protein (ALS5) gene, ALS5-1
ACCESSION AF068866
VERSION AF068866.1 GI:4903268
KEYWORDS
SOURCE Candida albicans.
ORGANISM Candida albicans.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 3813)
AUTHORS Hoyer, L.L. and Hecht, J.E.
TITLE The ALS5 gene of Candida albicans and analysis of the Als5p
N-terminal domain
JOURNAL Yeast 18 (1), 49-60 (2001)
MEDLINE 21064501
PUBMED 11124701
REFERENCE 2 (bases 1 to 3813)
AUTHORS Hoyer, L.L., Ho, M. and Hecht, J.E.
TITLE The ALS5, ALS6 and ALS7 genes of Candida albicans
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 3813)
AUTHORS Hoyer, L.L., Ho, M. and Hecht, J.E.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1998) Veterinary Pathobiology, University of
Illinois at Urbana-Champaign, 2001 S. Lincoln Avenue, Urbana, IL
61802, USA

FEATURES
source Location/Qualifiers
1. 3813
/organism="Candida albicans"
/strain="1161"
/db_xref="taxon:5476"
/chromosome="3"
/note="SfiI fragment p"
gene 1. 3813
/gene="ALS5"
/allele="ALS5-1"
1. 3813
/gene="ALS5"
/codon_start=1
/transl_table=12
/product="agglutinin-like protein"
/protein_id="AAD32849.1"
/db_xref="GI:4903269"
translation="MIOQFTLLFLYLSFATAKATGIENSIDSLTWSNAGNVPKPG
YPTNAVUGNSLDTSANPGDTFILNMFVKFTASQKSVDLDTAGVKYATCQFVS
EFTFTSLTCTVNDALKSSIAKFTVTLPIAFNVGGTGSIDLSKCKFTAGINTVF

NDGSKLSIAVNFPEKSTVDKSGYLLTTRMPMSLNKIALYVAPQCEGYTSGTGMGFST
 SYGDAVDCNSHIGISKYNDVNDWHPVTSFYSFKSCSPGSIITVQNPVAPRPF
 DTSPTAVTLLPFPNPSVDKTKTIELQLPITPTITTSYGVGTSYSTAPIGETATILI
 VDPYHTTIVTSKWTGIITMTTRITPTIDIVVVOVPLPNTTITITQWSESFTS
 TITLNKPEGTVSVIVKEPHNPTVTTFEFWSEYATETITPLGSDSIVIHDPLEE
 SSSPTAIESDSSNLSAQSSESSVEQSSSIVGLSSSDIPLSSDMPSSSGTUSSE
 SSTVSSDSSSSSELSSTPSSSSSISDITNFDWSSSSDLESTISWSSSIDAQ
 SSSQVSVNSISPTQSTSSGSESTSVTDILVSDASSILNDSILSSYPSPTISL
 SDOPPHIAGEPOSRSSTIASSTVEISDVLVSLTSDPTSSPDSSSILNDSPPSD
 ESDIASSSSTLVAPSFSSSSSSLSLIPHYVNSTYHASESESSSVSPVASES
 ANDHTLSTEDTSTSLIGDSSDSTFCRRDNGDCIVGTITSSIDSEOTSDVTTTS
 SPVASTTSAEOSTDIPNDISSTQTSASSTKSSVSVDVTVNLSILSETSTLSSD
 GTSDDTSISSTNDGTINAGSHKSTASTKSSIOKTGVTLSSYSLTKLSSSTDI
 TIELITELATIEDNEPTFTSPSSHSELSFSSDSSVLSKQVDRRESTIKTSPSTDTVT
 VSSLVHSTPASTATLGENSFNVASPLNATSLRSTSSSNHATSESSGVKSEASV
 EAPSPSTSDNRLSYSTEBAREGITYANSSTNNLITESQVAAFTDSTVLLENLYVT
 STFDNNSAAVDQPSKTKSIEESIMNPDSTNETNNGFIATLSQAQVSSSIHSELST
 TTKATTDASNMGDSAAANSOPTTLIQVATSSYNOPILITTYAGSSSATKHPSWLLKFI
 SVALFFFL*

BASE COUNT 1172 a 819 c 641 g 1181 t

ORIGIN

Query Match 31.4%; Score 1188.8; DB 8; Length 3813;
 Best Local Similarity 78.8%; Pred. No. 7.9e-205;
 Matches 1418; Conservative 0; Mismatches 382; Indels 0; Gaps 0;

QY 1 ATGCTTCAACAATTTACATGTTTATCTTATATTTGTCATTTGCAATGCAAGTGCAAGACATC 60
 DB 1 ATGATTTCAACAATTTACATGTTTATCTTATATTTGTCATTTGCGACTGCAAGGCGATC 60

QY 61 ACTGGTGTGTTTGTAGTTTAAATTCATTAACCTTGGTCCCAATGCTGCTAAATATGCTTTC 120
 DB 61 ACTGGTGTGTTTGTAGTTTAAATTCATTAACCTTGGTCCCAATGCTGCTAAATATGCTTTC 120

QY 121 AAAGGCGAGGATACCAACTTGGAAATGCTGTTGGGTGGTCTTAGATGGTACCAGT 180
 DB 121 AAAGGCGAGGATACCAACTTGGAAATGCTGTTGGGTGGTCTTAGATGGTACCAGT 180

QY 181 GCCAATCAGGGATACATTCACATGTAATATGCATGCTGTGTTAAATATACATCTCA 240
 DB 181 GCCAATCAGGGATACATTCACATGTAATATGCATGCTGTGTTAAATATACATCTCA 240

QY 241 CAAACATCTGTGTTAACTGCGGATGGTGTAAATATGCTACTGTGCAATTTATCT 300
 DB 241 CAAACATCTGTGTTAACTGCGGATGGTGTAAATATGCTACTGTGCAATTTATCT 300

QY 301 GGTGAAGAAATTCACAACTTTTCTACATTAACATGCTGTGAACGAGCGTTTGAATCA 360
 DB 301 GGTGAAGAAATTCACAACTTTTCTACATTAACATGCTGTGAACGAGCGTTTGAATCA 360

QY 361 TCCATTAAGGCATTTGGTACAGTTTACTTTACCAATTGCATTCATTTGGTGGAAACAGGT 420
 DB 361 TCCATTAAGGCATTTGGTACAGTTTACTTTACCAATTGCATTCATTTGGTGGAAACAGGT 420

QY 421 TCATCAACTGATTTGGAGATTTCTAAATGTTTTACTGCTGTGTACCAATFACAGTCACATTT 480
 DB 421 TCATCAACTGATTTGGAGATTTCTAAATGTTTTACTGCTGTGTACCAATFACAGTCACATTT 480

QY 481 AATGATGGTGAATAAGATATCTCAATTTGATGTTTGAAGTTTGAAGATCAACCGTTGATCCA 540
 DB 481 AATGATGGCAGTAAAAAGCTCTCAATTTGCTGTTAAATTTGAAAAAGTCAACAGTTGATCGA 540

QY 541 AGTCATATTTGATGCTTCCAGATTTATGCCAGTATGCTCAATTAAGTTCACAACTCTTTTT 600
 DB 541 AGTGGGTATTTGACTACTTCCAGATTTATGCCAGTATGCTCAATTAAGTTCACAACTCTTTTT 600

QY 601 GTGCGACCACAATCTGAAAAATGTTTACACATCTCGGTACAAATGGGTTCTCCAGTAGTAAC 660
 DB 601 GTGCGACCACAATCTGAAAAATGTTTACACATCTCGGTACAAATGGGTTCTCCAGTAGTAAC 660

QY 661 GGTGACGTGCTATTTGATGCTCAATATTTTCATATTTGGTATTCACAAAAAGGATTAATGAT 720
 DB 661 GGTGACGTGCTATTTGATGCTCAATATTTTCATATTTGGTATTCACAAAAAGGATTAATGAT 720

DB 661 GGGATGTTGCTATTTGACTGTTTCAAAATGTACATATTTGTTTCAAAAGGAGTAATGAT 720

QY 721 TGGAAATATCCGGTTTTCATCTCAATCATTTAGTTACATTAACATTTGTACATCTAATGA 780
 DB 721 TGGAAATATCCAGTTTACGCTCGAATCATTTAGTTACATTAACATTTGTTCATCTTTGTT 780

QY 781 ATTCAAGATTAATAATCAAAATGTACTGCTGCTTATCGTCCATTTATTTGATGCTTTATAT 840
 DB 781 ATCTCTATACATATCAAAATGTCTCTCCCGGTATCTGCCATTTATTTGACGCTTATAT 840

QY 841 TCTGCTACAGATTTAAACCAATATATCTTTAGCATATACCAATATATCTTTGCTGCTGCC 900
 DB 841 TCTCCTCAGATAATAATCAACAGTCAATTTGCTGTATAAAATAGCATATATCTTTGTTGAT 900

QY 901 AGTCGCTGCAAGATTAACCTTTACATTTAAGATGAGTGGATACAAAGATAGTATGCC 960
 DB 901 GATTATTGGCAACATGCACCTTTCACTTTAAATGAGTGGATATAAGAAATAGTATGCC 960

QY 961 GGATCTAACGGTATTTGTCATTTGTTGCTACAACTAGAACAGTTTACAGACAGTACCAGTCT 1020
 DB 961 GGATCTAACGGTATTTGTCATTTGTTGCTACAACTAGAACAGTTTACAGACAGTACCAGTCT 1020

QY 1021 GTCACTACTTTACCAATTCATCAAGTGTGATAAAACCAAAACAAATCGAAATTTGCAA 1080
 DB 1021 GTCACTACTTTACCAATTCATCAAGTGTGATAAAACCAAAACAAATCGAAATTTGCAA 1080

QY 1081 CCTATTCACCAACTACCATCACACTTCATATGTTGGTGTGAGTACTTCCCTATCTGACT 1140
 DB 1081 CCTATTCACCAACTACCATCACACTTCATATGTTGGTGTGAGTACTTCCCTATCTGACT 1140

QY 1141 AAGACTGCACCAATTTGGTGAACAGTACTTGTATTTGTTGATGTCCTCATATCATACTACC 1200
 DB 1141 AAGACTGCACCAATTTGGTGAACAGTACTTGTATTTGTTGATGTCCTCATATCATACTACC 1200

QY 1201 ACAACTGTTTACAGTGAATGAGACAGGAACAATCTACTACACACAACTCGTACCAATCCA 1260
 DB 1201 ACAACTGTTTACAGTGAATGAGACAGGAACAATCTACTACACACAACTCGTACCAATCCA 1260

QY 1261 ACTGATTCATTTGACACAGTGGTGTACAAGTTCACACTGCCAAATCCCAACTGTTAGTACT 1320
 DB 1261 ACTGATTCATTTGACACAGTGGTGTACAAGTTCACACTGCCAAATCCCAACTGTTAGTACT 1320

QY 1321 ACTGAATATTTGCTCTCAGTCTCTTGTCTACAACTACAGTTTACTGCTCTCCAGGTTGGT 1380
 DB 1321 ACCCAGTTTGGTCAAGTCTATTTACTAGTACTACTACAACTACCAACCAAGCCAGAGGC 1380

QY 1381 ACCGATCTGTGATTTATCAGAGAGCCCAACCAATCTACTGCTACTGCTACTGCTACTGGA 1440
 DB 1381 ACAGACTCAGTCAATTTGTCAGGAAACCAACAATCTCTGTTTACCACACAGAGTTTGG 1440

QY 1441 TCACAATCTTTGCTTACTTACTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1500
 DB 1441 TCAGATTCATATGCCACTACTGAAACAATTTACTACAGGCCACTTGGTACTGATAGTATC 1500

QY 1501 ATTATCAGAGAACACCAAAATCCAACTGCTACTACAAACCGAGTATTTGGTCTCAATCTTT 1560
 DB 1501 GTTATACATGATCCATTTGGAAGACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560

QY 1561 GCTACTACTACTACAGTTTACTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
 DB 1561 AATATTTCAAGCTCAGCTCAAGAAATCATCCAGTCTGTTGAGCAGTCTCAAGTATAGTT 1620

QY 1621 CCTCCAAACCAACTGTCACCACTCAATTAATTTGTTGCCAATCTTACGCAACCAACT 1680
 DB 1621 GGATTTGTCATCAAGTTTCAGATATACCAATTAAGTTTCAGACATGCCATCATCGAGCTCA 1680

QY 1681 ACTGTGACTGCTCTCCAGAGGCACTGACTCAGTAATTTATCAGAGAACCAACCAACCA 1740
 DB 1681 GGGTTAATCATCTAGTGTGCT 1740

QY 1741 ACTGTCTACTACTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1800
 DB 1741 AGTGAGTTTCTACATTTTCT 1800


```
RESULT 7
AR044076
LOCUS AR044076 1071 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5817466.
ACCESSION AR044076
VERSION AR044076.1 GI:5965541
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1071)
AUTHORS Hoyer,L.L., Livi,G.P. and Shatzman,A.R.
TITLE Conserved yeast nucleic acid sequences
JOURNAL Patent: US 5817466-A 1 06-OCT-1998;
FEATURES
Location/Qualifiers
source
1..1071
/organism="unknown"
BASE COUNT 317 a 322 c 159 g 273 t
ORIGIN

Query Match 28.2%; Score 1069.4; DB 6; Length 1071;
Best Local Similarity 99.9%; Pred. No. 3.8e-183;
Matches 1070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1300 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCCTTGGTACACCACTACA 1359
DB 1 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCCTTGGTACACCACTACA 60
QY 1360 GTTACTGCTCCCTCCAGGTGGTACCGATAGTGTGATTATCAGAGAGCCACCAACCACTACT 1419
DB 61 GTTACTGCTCCCTCCAGGTGGTACCGATAGTGTGATTATCAGAGAGCCACCAACCACTACT 120
QY 1420 GTCACACTACTGATATATTGGTCAAAATCCCTTGGTACTACTACTACTGTTACTGCTCT 1479
DB 121 GTCACACTACTGATATATTGGTCAAAATCCCTTGGTACTACTACTACTGTTACTGCTCT 180
QY 1480 CCAGGTGGTACTGACTCAGTAATATCAGAGAACCAACCAATCCAACTGTCACACCA 1539
DB 181 CCAGGTGGTACTGACTCAGTAATATCAGAGAACCAACCAATCCAACTGTCACACCA 240
QY 1540 GAGTATTGCTCAATCCTTGGTCTACTACTACTAGTACTGCTCCCTCCAGGTGGTACT 1599
DB 241 GAGTATTGCTCAATCCTTGGTCTACTACTACTAGTACTGCTCCCTCCAGGTGGTACT 300
QY 1600 GACTCAGTAATATCAGAGAACCTCCAAACCAACTGTCACCACTGATATTTGGTCC 1659
DB 301 GACTCAGTAATATCAGAGAACCTCCAAACCAACTGTCACCACTGATATTTGGTCC 360
QY 1660 CAATCTTAGCCAAACCACTACTGCTGCTCCTCCAGGAGGCACTGACTCAGTAAT 1719
DB 361 CAATCTTAGCCAAACCACTACTGCTGCTCCTCCAGGAGGCACTGACTCAGTAAT 420
QY 1720 ATCAGAGAACCAACCACTGTCACACTACTGATATGATATGCTGTCACATATATGCC 1779
DB 421 ATCAGAGAACCAACCACTGTCACACTACTGATATGATATGCTGTCACATATATGCC 480
QY 1780 ACCACTACCACTGTAATGTCACCACTGTCACCTGTCACCTGTTATCATTTAGAGAGCCA 1839
DB 481 ACCACTACCACTGTAATGTCACCACTGTCACCTGTTATCATTTAGAGAGCCA 540
QY 1840 CCAAAACCACTGTCACACTACTGATGATTTGGTCTCAATCCTTGGTACTACCACT 1899
DB 541 CCAAAACCACTGTCACACTACTGATGATTTGGTCTCAATCCTTGGTACTACCACT 600
QY 1900 GTAACCTGTCACCAAGTGGCACTGATAGTGTATCATTTAGGGAACCAACCACT 1959
DB 601 GTAACCTGTCACCAAGTGGCACTGATAGTGTATCATTTAGGGAACCAACCACT 660
QY 1960 GTCACCACTACTGATGCTCAATCATATGATCAACCACTACTACCACTTACCGTCCA 2019
DB 661 GTCACCACTACTGATGCTCAATCATATGATCAACCACTACTACCACTTACCGTCCA 720
```

```
2020 CCTGTGAAACTGATACCGTTCCTTATCAGAGAGCCACCAACCACTACTGTCACACTACT 2079
721 CCTGTGAAACTGATACCGTTCCTTATCAGAGAGCCACCAACCACTACTGTCACACTACT 780
2080 GAATACTGGTCTCAATATATATGCTCAACCACTGTTTACTGACCACTGTTGGTGAACCC 2139
781 GAATACTGGTCTCAATATATGCTCAACCACTGTTTACTGACCACTGTTGGTGAACCC 840
2140 GATACCGTTCCTTATCAGAGAGCCACCAACCACTACTGTCACACTACTGTAATCTGGTCT 2199
841 GATACCGTTCCTTATCAGAGAGCCACCAACCACTACTGTCACACTACTGTAATCTGGTCT 900
2200 CAATCATATGCTTACAAACCACTGTTTACTGACCACTGTTGGTGGTACCACTACTGTTATC 2259
901 CAATCATATGCTTACAAACCACTGTTTACTGACCACTGTTGGTGGTACCACTACTGTTATC 960
2260 ATTAGAGAGCCACCAACCACTACTACTACTGATATTTGGTCACAATCATTTGGC 2319
961 ATTAGAGAGCCACCAACCACTACTACTACTGATATTTGGTCACAATCATTTGGC 1020
2320 ACAACCACTGTTTACTGCTCCTCCAGGTGGTACTGACACTGTTGATTATC 2370
1021 ACAACCACTGTTTACTGCTCCTCCAGGTGGTACTGACACTGTTGATTATC 1071

RESULT 8
I65632.
LOCUS I65632 1071 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 1 from patent US 5668263.
ACCESSION I65632
VERSION I65632.1 GI:2482202
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 1071)
AUTHORS Hoyer,L.L., Livi,G.P. and Shatzman,A.R.
TITLE Conserved yeast nucleic acid sequences
JOURNAL Patent: US 5668263-A 1 16-SEP-1997;
FEATURES
Location/Qualifiers
source
1..1071
/organism="unknown"
BASE COUNT 317 a 322 c 159 g 273 t
ORIGIN

Query Match 28.2%; Score 1069.4; DB 6; Length 1071;
Best Local Similarity 99.9%; Pred. No. 3.8e-183;
Matches 1070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1300 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCCTTGGTACACCACTACA 1359
DB 1 CCAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCCTTGGTACACCACTACA 60
QY 1360 GTTACTGCTCCCTCCAGGTGGTACCGATAGTGTGATTATCAGAGAGCCACCAACCACTACT 1419
DB 61 GTTACTGCTCCCTCCAGGTGGTACCGATAGTGTGATTATCAGAGAGCCACCAACCACTACT 120
QY 1420 GTCACACTACTGATATATTGGTCAAAATCCCTTGGTACTACTACTACTGTTACTGCTCT 1479
DB 121 GTCACACTACTGATATATTGGTCAAAATCCCTTGGTACTACTACTACTGTTACTGCTCT 180
QY 1480 CCAGGTGGTACTGACTCAGTAATATCAGAGAACCAACCAATCCAACTGTCACACCA 1539
DB 181 CCAGGTGGTACTGACTCAGTAATATCAGAGAACCAACCAATCCAACTGTCACACCA 240
QY 1540 GAGTATTGCTCAATCCTTGGTCTACTACTACTAGTACTGCTCCCTCCAGGTGGTACT 1599
DB 241 GAGTATTGCTCAATCCTTGGTCTACTACTACTAGTACTGCTCCCTCCAGGTGGTACT 300
QY 1600 GACTCAGTAATATCAGAGAACCTCCAAACCAACTGTCACCACTGATATTTGGTCC 1659
DB 301 GACTCAGTAATATCAGAGAACCTCCAAACCAACTGTCACCACTGATATTTGGTCC 360
QY 1660 CAATCTTAGCCAAACCACTACTGCTGCTCCTCCAGGAGGCACTGACTCAGTAAT 1719
DB 361 CAATCTTAGCCAAACCACTACTGCTGCTCCTCCAGGAGGCACTGACTCAGTAAT 420
QY 1720 ATCAGAGAACCAACCACTGTCACACTACTGATATGATATGCTGTCACATATATGCC 1779
DB 421 ATCAGAGAACCAACCACTGTCACACTACTGATATGATATGCTGTCACATATATGCC 480
QY 1780 ACCACTACCACTGTAATGTCACCACTGTCACCTGTCACCTGTTATCATTTAGAGAGCCA 1839
DB 481 ACCACTACCACTGTAATGTCACCACTGTCACCTGTTATCATTTAGAGAGCCA 540
QY 1840 CCAAAACCACTGTCACACTACTGATGATTTGGTCTCAATCCTTGGTACTACCACT 1899
DB 541 CCAAAACCACTGTCACACTACTGATGATTTGGTCTCAATCCTTGGTACTACCACT 600
QY 1900 GTAACCTGTCACCAAGTGGCACTGATAGTGTATCATTTAGGGAACCAACCACT 1959
DB 601 GTAACCTGTCACCAAGTGGCACTGATAGTGTATCATTTAGGGAACCAACCACT 660
QY 1960 GTCACCACTACTGATGCTCAATCATATGATCAACCACTACTACCACTTACCGTCCA 2019
DB 661 GTCACCACTACTGATGCTCAATCATATGATCAACCACTACTACCACTTACCGTCCA 720
```

JOURNAL	Submitted (11-SEP-1997)	Veterinary Pathobiology, University of Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES	Location/Qualifiers	
source	1. .1404	
	/organism="Candida albicans"	
	/strain="1161"	
	/db_xref="taxon:5476"	
	/chromosome="6"	
	/map="6c"	
gene	1. .>1404	
	/gene="ALS2"	
	/note="agglutinin-like sequence"	
	/allele="ALS2-2"	
CDS	1. .>1404	
	/gene="ALS2"	
	/codon_start=1	
	/transl_table=12	
	/product="agglutinin-like protein"	
	/protein_id="AAC64237.1"	
	/db_xref="GI:3598675"	
	/translation="MLQFQLLLSLCVSAKATVITGVFNSDLSTRAGNYAYKGN RPTWALVGMISIDGTSANPGDFTFLNMPVKEFTIDQTSVDLTAEGVKYATCFYSGE ERTTSSLLKCVSNVLTSSIKALGTVLPISFNVGGTGSVDLSSOCFAGINTVTF LGDQKKISIDVDFKTNEDASGDFIASRLIPFINKVSIITYAPOCAAGNYTSGMAGFIV LVTGTTIDCSNVHVGITKGLNDWMPVSSDSLSYNKTCSSGTGISITYENVPAGYRPF DYYTSVSGNQRLRYTNDYACVGSGLASQKPNLRGLYNNSEANSNGFVIAVTRVTV DSTTATLTPNPSVDYKTKTEILQPIPTTTITTSYGVVTTSYSTKTPAIGETATVIV DVPYHTTPTVTSEWNGTITATITRTNPDSIDTVVQVQVSPNPVSTTEVWSQSYATT TTVTAPPGGGTDSVIIREP"	
repeat_region	1297. .1404	
	/note="first of multiple copies of 108bp imperfect repeat"	
	/rpt_type=tandem	
BASE COUNT	421 a 266 c 255 g 462 t	
ORIGIN		
Query Match	26.9%;	Score 1016.6; DB 8; Length 1404;
Best Local Similarity	83.2%;	Pred. NO. 1.2e-173;
Matches 1170;	Conservative	0; Mismatches 234; Indels 3; Gaps 1;
1	ATGCTTCAACAATTTACATTTGTTATTCCTATATTTGTCATTTGCAAGTGCAGACCAATC 60	
1	ATGGCTTTTCAAAATTTTGTGGCTGAAGCCTCTGTGTATCAGTTGCTACTGCAAAAGATTAT 60	
61	ACTGGTGTGTTTTGATAGTTTTAAATTCATTAACTTGGTCCAATGCTGCTCAATATGCTTTC 120	
61	ACGGGTGTTTTCAATAGTTTTCATTCGTTGACATGGCAAGAGCTGGTAATATGCTTAT 120	
121	AAAGGCCAGGATACCCAACTTGGAAATGCTGTTTTGGTGTGGTCTTAGATGGTACCACT 180	
121	AAGGGCCCAATAGACCAACTTGGAAATGCTGTTTTGGGCTGCTTTAGATGGTACTAGT 180	
181	GCCATCCAGGGGATACATTCACATTTGAATATGCCATGTGTTTTAAATATATACATCTCA 240	
181	GCAAAATCCAGGAGACACATTCACATTTGAATATGCCATGTGTTTTAAATTTATACCGAT 240	
241	CAAAATCTGTTGATTTAACTCCCGATGGTGTAAATATGCTACTGTGCAATTTTATTCT 300	
241	CAAAATCTGTTGATTTGCTGCTGAAGGTGTTAAATATGCTATGCTACGCTTTTATTC 300	
301	GGTGAAGAAATCCAAACTTTTTCTACATTAACATGTACTGTGAACGAGCTTTGAAATCA 360	
301	GGTGAAGAAATTTACAACATTTTCTTCATTAATAATGCTGTGAGCAATACTTTAAATCA 360	
361	TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTCGATTCATATGTTGGTGAACAGGT 420	
361	TCTATTAAGGCCTTTGGGTACGGTACTTTTACCAATTTCAATTTAATGTTGGTGAACAGGT 420	
421	TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGTGATACCAATACAGTACACATTT 480	
421	TCATCGGTTGATTTGGAAGATTCTCAATGTTTTAAGGCTGGCACCACACAGTTACTTTT 480	
481	AATGATGGTGATAAAGATATCTCAATTTGATGTTGAGTTTGAAGAGTCAACCGTTTATCC 540	

ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE 1 (bases 1 to 1404)
AUTHORS Hoyer, L.L., Payne, R.L. and Hecht, J.E.
TITLE Identification of Candida albicans ALS2 and ALS4 and localization
of als proteins to the fungal cell surface
J. Bacteriol. 180 (20), 5334-5343 (1998)

JOURNAL 98440424
MEDLINE
PUBMED 9765564

REFERENCE 2 (bases 1 to 1404)
AUTHORS Hoyer, L.L.
TITLE Direct Submission
JOURNAL Submitted (11-SEP-1997) Veterinary Pathobiology, University of
Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA

FEATURES
source Location/Qualifiers
1..1404
/organism="Candida albicans"
/strain="1161"
/db_xref="taxon:5476"
/chromosome="6"
/map="6; C"
1..>1404
/gene="ALS2"
/codon_start=1
/transl_table=12
/product="agglutinin-like protein"
/protein_id="AAC64235.1"
/db_xref="GI:3598670"
/translation="MLLPGLLSLCVSATAKVIITGVNFSFDSLWTRAGNVAYKGNP
RPTNVAQWGLSDGTSANPDGFTLNMPCKFKPITDOTSVDLTAEVKVATCGYGFGE
EFTTFSLACTVSNLTSSIKALTGTVLPISFNVGDTGSSVDLESQCFKAGNTVTFE
NQGDKKSIDVDFEKTNEASGYFIASRLIPSKVSIYVAPQCAAGTYTGAMGRFLV
LFGDTIDCSNVRHLYTGLNDWNPVPSDSLSYNKTSSTGISIYENVPAGYRPF
DVTYSVGNQGLRYITNDYACVGSLSQKPNRLRNGNNSANGFVIVATRTV
DSTAVTITLFPNPSVDKTIIEILQIPTTTTXYGVTTSYSTKTAIGETATVIV
DVPYHTTNTVTSMTGTCTITTTTTRTNTDSDIDTVVQVPSNPPTVTTEVWSOYAT
TIVTAPPGTGDSVLIIEP"
repeat_region 1297..1404
/note="first of multiple copies of imperfect 108 bp
repeat"

BASE COUNT 422 a 268 c 253 g 461 t
ORIGIN

Query Match 26.6%; Score 1006; DB 8; Length 1404;
Best Local Similarity 82.7%; Pred. No. 9.8e-172;
Matches 1163; Conservative 0; Mismatches 240; Indels 3; Gaps 1;

QY 1 ATGCTTTCAACAATTTACATTTGTTATTCTTATTTCTCAATTCGAATGCGAAAGACAAT 60
Db 1 ATGCTTTTACAATTTTGTGCTAACCCCTCTGTGATCAGTTCCTACTGCGAAAGTTATT 60

QY 61 ACTGGTGTTTTTGATAGTTTTTAATTCATTAACTTGGTCCAAATGCTGCTAATTATGCTTTC 120
Db 61 ACGGGTGTTTCAATAGTTTGTGATTCGTTGACATGGACAAGAGCTGGTAATTATGCTTAT 120

QY 121 AAGGGCCAGGATACCCAACTTGGATGCTGTTTTGGGTGGTTCCTAGATGTCACAGT 180
Db 121 AAGGGCCAAATAGACCAACTTGGAAATGCTGTTTTGGGCTGGTCTTTAGATGGTACTAGT 180

QY 181 GCCAATTCAGGGGATACATTTACATTTGAATATGCCATGTGTGTTTAAATATATACTTCA 240
Db 181 GCAATTCAGGAGACACATTCACATTTGAATATGCCATGTGTTTTTAAATTTATACCGAT 240

QY 241 CAAACATCTGTTGATTTAACTGCGCGATGGTGTGTTTAAATATGCTACTTGTCAATTTTATCT 300
Db 241 CAAACATCTGTTGATTTGACTGCTGAAGGTGTTAAATATCCTACATGTCAGTTTTATTCA 300

QY 301 GGTGAAGAATTCACAACCTTTTCTACATTTAAACATGACTGTGGAACGACGCTTTGAAATCA 360
Db 301 GGTGAAGAATTTACAACATTTTCTTCATTTAAATGTACTGTGAGCAATACCTTTAATCATCA 360

Db	26	TTTTCTCTCTATTGTCACGATACGAATGGCAAAACATATATCGGAGTTTTTCACGAGTTTCA	85
QY	83	ATTCATTAACTGGTCCAAATGCTGCTAAATATATGCTTTCAAAGGGCCAGGATACCCAAT	142
Db	86	ACTCATGACCTATATCTACTGTAATCTACCAATATGCGGGTCTGTTATCCAACT	145
QY	143	GGAACTGCTGTTTGGGTTGGCTTTAGATGTTACAGTGGCAATCCAGGGATACATTTCA	202
Db	146	GGACTGCTGTTTGGGTTGGAGCTTGGACGGAACACTAGCTAGTCCAGGTGATACATTTA	205
QY	203	CATTGAATATGCCATGCTGTTTAATATATCTACTTTCACAACTCTGTTGTTAACTG	262
Db	206	CATTGGTGCCTGCTGCTTTTCAAATTTATTCACACAAACTTTCAGTAGACTTTAACTG	265
QY	263	CGATGGTGTAAATATGCTACTTGTCAATTTATTTCTGGTGAAGAAATTCACAACTTTT	322
Db	266	CTAATGGTGCATGATGCACATGCTACTTTCCATGGAGGGAAGACTTTACTACTTTT	325
QY	323	CTACATTAACATGCTACTGTGAACGAGCTTTGAAATCATCCATTAAGGCATTTGGTACAG	382
Db	326	CAAGTATGATGCTGTAGTAAATATGGCTATCTTCAAATATCAGAGCGTTTGGTACCG	385
QY	383	TTACTTTACCAATTCGATTCATGTTGGTGGACAGGTTTCATCACTGATTTGGAGATT	442
Db	386	TCAGGCTACCAATTTCAATTCGATGTTGGTGGAACTGGTTCATCTCTCAACATTCAGATT	445
QY	443	CTAAATGTTTTACTGCTGTACCAATACAGTCACATTTAAATGATGTCGATAAAGATATCT	502
Db	446	CAAAGTGTTCCTGCTGGACGACACTGTAAACATTTACAGACGGCATCCAAAATTT	505
QY	503	CAATTTGATGTTGAGTTTGAAGTCAACCGTTTGATCCAAAGTCATATTTGATGTTCCA	562
Db	506	CTACTACAGTCAATTTCCCTAAGACTCCACAACTATCTAGTAGCTTGGTTTATTTCGCAA	565
QY	563	GAGTTATGCCAGTCTCAATAGGTCACAACTCTTTTGTGGCACCACCAATGTGAAATG	622
Db	566	GGGTTATTCCAAAGTCTTGATAAATTTATCTAGCTTGTGTTGCTCTCAGTGTACTGCTG	625
QY	623	GTACACATCTGGTACAAATGGGTTCTCCAGTAGTAACGGTGCACGTTGCTATTGATTTCT	682
Db	626	GATATGCATCCGGTGTGCTCGATTTTCAGCACAAGAGATGATGTGACAAATGATTTGT	685
QY	683	CAATATTCATATGTTGATACAAAAGGATTAATGATTTGGAATTTATCCGGTTTCATCTG	742
Db	686	CTACTATACATGTGGGAATAACAAATGGTTTGAATAGTTGGAATATGCCAGTATCATCAG	745
QY	743	ATCATTTAGTTACACTAAACTCTACATCTAATGGAATTCAGATTAATATCAAAATG	802
Db	746	AATCATTTTCTTACACCAAACTTGTACACCAACAGTTTTTATTTACTTATGAAATG	805
QY	803	TACCTGCTGTTATCGTCCATTTATTTGATGCTT-----ATATTTCTGTACAGATGTTA	856
Db	806	ITCCTGCAAGTTATCGTCCATTTATTTGATTTCTTACGTGNAAAAATCAGCAACGCAACA	865
QY	857	ACCAATATCTTTAGCATATACCAATGATTTATCTGTTGGCGAGTCTGTGCAAAATG	916
Db	866	ATGGATTAAATTTGAATTTACAGCAATATATCAAAATTTGATGGATGGCAAAAAGGAAATG	925
QY	917	RACCTTTCACTTTAGATGGATGCTGATACAGAAATAGTATGATCCGGATCTACGCTATG	976
Db	926	ATCCTCTTATATACCTTTTGGACATCATACAAATATGATGTCAGGATCCCAATGGAGCTG	985
QY	977	TCATTTGGTGTACAACTAGAACAGTTACAGACAGTACCACCTGCTGTCTACTTTTACCAT	1036
Db	986	CGGTAGTTGTTACTAGGACAGTCACTGATTTCTACAACAGCAATTAACCAATTTACCGT	1045
QY	1037	TCAAATCCAGTGTGATAAAACCAAAACAAATGGAATTTTGAACCTATTTCCAACTA	1096
Db	1046	ITGATCCCAAGTTTATGATAAAACCAAAACCAATTTGAAGTAAATAGAACCCCTACTACCA	1105
QY	1097	CCATCACAACTTCATATGTTGGTGTGACTACTTCTCTATCTGACTGAAGACTGCACAAATG	1156
Db	1106	CTATTACTACTTCATATGTTGGATTTCTTACTTCTACTTTCTCAGGAAGACTGCAACTATG	1165
QY	1157	GTAAACACGCTACTGTTTATTTGTTGATGTCATATCATCTACTACCAACTGTTTACCAGTG	1216
Db	1166	GAGGAACGCACTGTTGTTGTTGATGTTCCCTATCATCACTACTACCCTACTACTAGTA	1225
QY	1217	AATGACAGGAAACAACTACTACCAACAACTCGTACCAATCGCAACTGATCAATTTGACA	1276
Db	1226	TATGACTGGATCAGCTACCACTCAAGTACTTATACAAATCCCACTGACTCGATTGATA	1285
QY	1277	CAGTGGTGGTCAAGTTCCTCACTGCCAAATCCAACTGTTAGTACTACTGAATTTGGTCTC	1336
Db	1286	CAGTTGTTGTACAAGTTCCTCACTGCCAAATCCAACTGTTACAACTACTCAGTTTGGTCAG	1345
QY	1337	AGTCTCTTGTCTACAACTACTACAGTTACTGCTCTCCAGGTCCTCCAGGTCCTACCTACTGTA	1396
Db	1346	GAAAGTGTCCCAACACCGAACTGTGACCACTGGACCAACAGGAGGATAGTGATCA	1405
QY	1397	TCAGAGAGCCCAAAACCACTACTGTCTACTACTGTGAAATATTTGGTGCATAATCTTTGCTA	1456
Db	1406	TCAAGGAGCCCAACAACTACTGTGACTTACCCTGAGTTTGGTGCAGAAATCATTTGCTA	1465
QY	1457	CTACTACTACTGTTACTGCTCTCCAGGTTGGTACTGACTCAGTAAATATTCAGAGAACCA	1516
Db	1466	CTACTGAGACAGTCAACCAATCCGAGGCACTGATGTTGATCATCAAGGAACCA	1525
QY	1517	CAATCCAACTGTCTACTACCAACCGAGTATTTGGTCTCTCAATCTTGTCTACTACTACTACAG	1576
Db	1526	ACAATCTCTACTGTTTACCACCACCGAGTTTGGTGCAGAAATCATTTGCTACTACTGAGACAG	1585
QY	1577	TTACTGCTCTCCAGGTTGTTACTGACTCAGTAAATTTATCAGAGAACTCCAAACCACTG	1636
Db	1586	TCACCAACTATCCAGAGGGAAGTACTGACTCAGTCATTTGTTAGAGAACCACAAATCCA	1645
QY	1637	TCACCACTACTGAATTTGGTCCCAATCTTACGCAACCACTACTGCTGACTGCTCTC	1696
Db	1646	TAAACAACCAACCGAGTTTGGTGCAGAAATCATTTGCTACTGAGACAGTCACCAACTATC	1705
QY	1697	CAGGAGGCACTGACTCAGTAAATTTATCAGAAACCAACCACTGCTGCTACTACTACTG	1756
Db	1706	CAGAGGGAAGTACTGCTCAGTCATTTAGGGAACCAACAACTCTACTGTTACCACCA	1765
QY	1757	AATCTGCTCAATATATGCGACACTTACCACTGTAATCTGTAACGCAACAGGTTGATG	1816
Db	1766	AGTTTGGTGCAGAGTCATTTGCTACTGAGACCATCACCACCTATCCAGAGGGA	1825
QY	1817	ACACTGTTATCATGAGAGCCCAACCACTGCTACACTACTGCTGAGTATTTGCTCTC	1876
Db	1826	ACTCAGTCATTTAGGGAACCAACCAATCCAACTGTAACAAACCCAGTTTGGTCAG	1885
QY	1877	AATCGTTTCTACTACCACTGTAATCTGTTCCCAAGTGGCACTGATCTGTTATCA	1936
Db	1886	AATCATTTGCTACTACTGAAACGATACTACAGGGCCCACTTGGCACTGATAGTATCGTTA	1945
QY	1937	TTAGGGAACCA	1947
Db	1946	TACATGATCCA	1956
RESULT 12			
AF201684			
LOCUS		6897 bp	DNA linear
DEFINITION			Candida albicans agglutinin-like protein Als7p (ALS7) gene,
			complete cds.
ACCESSION			AF201684
VERSION			AF201684.1
KEYWORDS			GI:9754770
SOURCE			Candida albicans.
ORGANISM			Candida albicans.
REFERENCE			Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS			Saccharomycetales; mitosporic Saccharomycetales; Candida.
TITLE			1 (bases 1 to 6897)
			Hoyer, L.L. and Hecht, J.E.
			The ALS6 and ALS7 genes of Candida albicans

[illegible][illegible]

Qy	964	TCTAACGGTATGTCAATGTTGGTCTCAACACTAGACAGACTTACAGACAGTAGTACCACCTGCTGTC	1023
Db	970	GCTATTGGTGCGAGACCTAGTTCGTACCACATCCACACTGATAGTTCCATTCACGAACT	1029
Qy	1024	ACTACTTTACCATCAATCCAAAGTGTGATAAACCAAAACAATCGNAATTTTGCACACT	1083
Db	1030	ACCACATTAACCTTCATTTCCGACTCCGAAACCAAAACAATTCCTAGTCTTTAGAGCCC	1089
Qy	1084	ATTCCAACCACTACCATCAACATTCATATGTTGGTGTGACTACTTCCCTATCTGACTAAG	1143
Db	1090	ATACCCACCACTAGCGTAAACAATTCACACCACTGGCTTGTGATCTGGTATTAATATAAG	1149
Qy	1144	ACTGCCACCAATTTGGTGAACAGCTACTGTATTGTTGATGTGCCATATACATAACACA	1203
Db	1150	AAAGCCACCAATTTGGTGACACAGCTACTGTTTCATTGATGTTCCACACACACAGTACT	1209
Qy	1204	ACTGTTTACCAGTGAATGGACAGGAACAATCACTACCAACCAACTCGTACCAATCCAAC	1263
Db	1210	ACTTTGACACATATTGGCAAGAATCAAGTACAGCGACAACCACTTACTTCGATGACAVA	1269
Qy	1264	GATTCAAATTGACACAGTGGGTACAGTTCCTGCTCCTCAGGTGGTACTGACTCAGTAAT	1323
Db	1270	GACTTGGTGATACTGTCAATGTGAAATTTCCATATCCCAATCCGACTATTATPAACA	1329
Qy	1324	GAATPATTTGGTCTCAGTCTTGGTACAAACCACTACAGTTACTGCTCCTCAGGTGGTACC	1383
Db	1330	CAATTTTGGTCAGGTAATAATTTAACTACTGACACACACAAAGAACCACTCTCGGTACT	1389
Qy	1384	GATACTGTGATTTATCAGAGAGCCACCAACCACTACTGTCTACTACTGAATATTTGGTCA	1443
Db	1390	GATAGTGTGATCATCAAGGAACCAACCAACCCCTACTGTGACAACGACCGAGTTTGGTCA	1449
Qy	1444	CAATCCCTTCTCTACTACTACTGTGTACTGCTCTCCAGGTGGTACTGACTCAGTAAT	1503
Db	1450	GAATCATTTGCCACTACTGAGACCAATCAACCACTATCCAGAGGGCACTGACTCAGTCA	1509
Qy	1504	ATCAGAAACCAACCAATCCAACTGTCACTTACAAACGAGTATTGGTCTCAATCCTTTGCT	1563
Db	1510	GTTAGGAACCAACCAACCCCTACTGTGACAACGACCGAGTTTGGTCAGANTCATTTGCC	1569
Qy	1564	ACTACTACTACAGTTACTGCTCTCCAGGTGGTACTGACTCAGTAATTTATCAGAGAACT	1623
Db	1570	ACTACTGAGACCAATCAACCGGTCAGAGGCACTGACTCAGTCAATTTGTAGGAACCA	1629
Qy	1624	CCAAACCAACTGTCAACCACTGAAATTTGGTCCCAATCTTACCAACCAACCACTACT	1683
Db	1630	CATAATCCAAGTGTGACAACAACCGAGTTTGGTCAGAAATCATTTGCCACTACTGAGAC	1689
Qy	1684	GTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTTATCAGAGAACCAACCAACCACT	1743
Db	1690	ATCACAACCGTCCAGAAGGCACTGACTCAGTCATTGTTAGGAACCAACCAATCCA	1749
Qy	1744	GTCACTACTACTGAATCTGGTCACAATCATATGCCAACACTTACCAGTGTAACTGCAACA	1803
Db	1750	GTGACAACAACCGAGTTTGGTCAGAAATCATTTGCCACTACTGAGACCACTCAACCGT	1809
Qy	1804	CCAGGTGGTACTGACACTGTTATCATTTAGAGAGCCACCAACCACTGCTACTACTACT	1863
Db	1810	CGAAGGCACTGACTCAGTCATGTTAGGAACCACTAATCCAACTGTGACACCAACG	1869
Qy	1864	GAGTATTTGGTCAAGTTCATTTGGCCACTACTGAGACCACTCAACCGTCCGAAGGCACT	1923
Db	1870	GAGTATTTGGTCAAGTTCATTTGGCCACTACTGAGACCACTCAACCGTCCGAAGGCACT	1929
Qy	1924	GATACTGTTATCATTTAGGAACCAACCAACCACTGTCCACCACTACTGAATACTGGTCT	1983
Db	1930	GATAGTGTGATCATCAAGGAACCAACCAACCCCTACTGTGACTTACCACCAAGTTTGGTCA	1989
Qy	1984	CAATCATATGCAACCACTACTACATTTACCGTCCCACTGTTGAACACTGATACCGTCTT	2043
Db	1990	GAATCATTTGCCACTACTGAGACATCAACCACTATCCGAAGGCACTGACTCAGTCA	2049
Qy	2044	ATCAGAGAGCCCAACCACTACTGTCACTACTGTAATAGTGGTCTCAATCATATGCT	2103

Db	2050	GTTAGGGAACCA	TAATCCAACTG	TGACAAACCGAG	TTTGGTCAGAA	TCATTTTGCC	2109		
QY	2104	ACAACACCACT	GTGTACTGCA	CCACCACTGGT	GAAACCGATAC	CGGTCTTATC	CAGAGACCA	2163	
Db	2110	ACTACTGAGAC	CAATCACCACG	GTCAGAGGCACT	GACTCAGTCAT	TGTTAGGAACCA	2169		
QY	2164	CCAAACCACT	GTCTACTACT	GAACTACTGGT	CTCAATCATATG	TCAACAACCACT	2223		
Db	2170	CACAATCCA	CTGTGACAA	CAACCGAGTTT	GGTCAGAACT	CAATTTGCTACT	CTAGAGACC	2229	
QY	2224	GTTACTGCAC	CAACCAAGTGG	TACCGATACTG	TTATCATTTAG	AGAGCAACCA	AAATCCAACA	2283	
Db	2230	ATCACCAACT	ATCCAGAGGA	CACTGACTCAG	TCAATTTGTAG	AGAAACCAACA	CAATCCAAC	2289	
QY	2284	GTTACTACT	CTGAATATGG	TGCACATCAT	TTTCCCAACCA	CCACCACAGT	TACTGCTCCT	2343	
Db	2290	GTACACAAC	CCGAGTTTGG	TGCAGAGTC	ATTTGCTACT	ACTGAGACAG	CAACCACTAT	2349	
QY	2344	CCAGTGGT	ACTGACACT	GTGATTTAT	-----CTAT	GAAACATGT	CAAGTTCA	2391	
Db	2350	CCAGAGGC	ACTGACTCAG	TCTCAGTCAT	TTGTAGGA	CAACCAACA	CAATCCA	CTGTGACAC	2409
QY	2392	AAGATTTCT	ACATCCTCCA	ATGATTAAC	CAGTATCAT	TCCATCAT	TTTTCCC	GTCCCTCAT	2451
Db	2410	GAGTTTTGG	TGCAGAACT	ATTTGTCACT	ACTGAAACG	ATAACTAC	GAGGCCAC	TTGGCACT	2469
QY	2452	TATGTCAC	AGCAACCTCC	GATTTTCTCA	CACTTTGAA	TTTCATCC	ATGAACTACT	CTCT	2511
Db	2470	GATAGTAT	PCGTTATAC	ATGATPCCA	TGGAAGAA	CTGCTTCT	AGTCTGCT	ATTTAGTGCA	2529
QY	2512	ACTTCTAT	CACTAGTGT	GTATGTTGT	GTCTTCTACA	CACTTTGGT	TACTGAAT	CAGAA	2571
Db	2530	AGTGATCT	TAATATTTCA	AGCTCAGCT	CAAGATCAT	CAGTCTGGT	TGAAACAGT	CAATTT	2589
QY	2572	ACAACAC	AGAACTGAT	TTGCGATG	TGGTAAAG	AGATGTTCT	AGATTTGCC	AGTCTCTCT	2631
Db	2590	ACTTCTG	TGCACGAG	ACTTCAAG	TATAGTTGA	ATTTGTCA	TCAAGATC	AGACATCC	2649
QY	2632	GGTATTG	TCACAAAT	CCAGATAG	CAATGAAT	CTCAATCGT	AACTAGTACT	GTTCTCTACT	2691
Db	2650	AGCTCAAT	TGGGTTAA	CACTCTAG	TGAGTCTGCT	CTCAAGT	TATATAG	ACTACTCT	2709
QY	2692	GCAAGT	ACAATGT	CTGATTC	CTTCAACT	GATGGTAT	TAGTGCTAC	ATCTTCTGAT	2751
Db	2710	TCAAGT	ACTAGCG	GAATCAT	CTATTTGCT	TCAAGTTAT	GATAGCTAT	TTTCGTC	2769
QY	2752	AATGTTT	CAAAATC	AGGAGTAT	CACTGTTACA				2781
Db	2770	GAGTCG	TCTACT	TTTATCTAG	TTCCGATAGA				2799

RESULT 13

ANALYSIS	1407 bp	DNA	linear	PLN 16-OCT-1998
LOCUS	Candida albicans agglutinin-like protein (ALS4) gene, 5' partial cds.			
DEFINITION	Candida albicans agglutinin-like protein (ALS4) gene, 5' partial cds.			
ACCESSION	AF024584			
VERSION	AF024584.1	GI:3598677		
KEYWORDS	1 of 2			
SEGMENT	Candida albicans.			
SOURCE	Candida albicans.			
ORGANISM	Candida albicans.			
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.			
AUTHORS	Hoyer, L.L., Payne, T.L. and Hecht, J.E.			
TITLE	Identification of Candida albicans ALS2 and ALS4 and localization of als proteins to the fungal cell surface			
JOURNAL	J. Bacteriol. 180 (20), 5334-5343 (1998)			
MEDLINE	98440424			
PUBMED	9765564			
REFERENCE	2 (bases 1 to 1407)			

AUTHORS Hoyer, L.L.
 TITLE Direct Submission
 JOURNAL Submitted (11-SEP-1997) Veterinary Pathobiology, University of Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA

FEATURES
 source Location/Qualifiers
 1..1407
 /organism="Candida albicans"
 /strain="1161"
 /db_xref="taxon:5476"
 /chromosome="6"
 /map="6C"
 1..>1407
 /gene="ALS4"
 /note="agglutinin-like sequence"
 /allele="ALS4-1"
 1..>1407
 /gene="ALS4"
 /codon_start=1
 /transl_table=12
 /product="agglutinin-like protein"
 /protein_id="AAC64239.1"
 /db_xref="GI:3598680"
 /translation="MLQFLLLSLCVSVATAKVITGVFNSFNSLITWANAASPYRGPA
 TPTVAVIGSLDGTASAGDTFTLDMPVFKFITDQISDLVADGRYATCNLSAE
 EFTTFSSVSCVTMTMTADTKAIGTFLPFSSVSGSDVDLANSQCFTAGINTVF
 NDGTSISTTVDFEKSTVASSDRILLRILPSLSQAVNLELPQBCANGTSGTMGFST
 AGTCATIDCSVHVIGISGLNDMNPISSEFSYTKTCTSTSLVTFONVPAGYRPFV
 DAYISATRVSSYTHQYINLACVGAASVDDSFTHWBGYSNSQAGSNGITIVVTRTV
 TDSITAVTILFPNSDTHKTIIEILOPIPTTIITSYGVVTSYSTKAPIGETATVI
 VDVPHTTIVTSWMTGTITTTTNTPTSDIDIVVQVPSNPFTVTTEYWSOSAT
 TTTVAPPGGDVSVIIEP"
 1300..1407
 repeat_region 391 a 294 c 253 g 469 t
 /note="first of multiple copies of 108bp imperfect repeat"
 /rpt_type=tandem

BASE COUNT 391 a 294 c 253 g 469 t

ORIGIN

Query Match 22.3%; Score 842.8; DB 8; Length 1407;
 Best Local Similarity 75.0%; Pred. No. 2.8e-142;
 Matches 1054; Conservative 0; Mismatches 352; Indels 0; Gaps 0;

QY 1 ATGCTTCAACAATTTACATTTGTTATCTCTATATTTGTCATTTGCAATGGCAAGTGCACAAACAATC 60
 Db 1 ATGCTTTTACAAATTTTCTGCTAAGCCCTCTGTGTATCAGTTGCTACGGCAAAAGTTATT 60

QY 61 ACTGCTGTTTGTAGTATTTAATTCATTAATCTGGTCCCAATGCTGCTTAATATGCTTTC 120
 Db 61 ACAGGTGTTTCAATAGTTTAAATCGTTAACTTGGGCCAATGCTGCTTCTTATCCATAT 120

QY 121 AAAGGCCAGGATACCAATCGTAAATGCTGTTGGTGTGCTTGTAGTGGTACCAAGT 180
 Db 121 AGAGTCCAGTACTCTCTACTTGGACCCGCTGTAATAGGATGGTCTTTAGATGGAGCTACT 180

QY 181 GCCAATCAGGGGATACATTCACATTTGAATATGCCATGCTGTTTAAATATATCTACTTCA 240
 Db 181 GCTAGTGTGGTGACACATTCACGTTGACATGCTGCTGTTTTCACAAATTTATTAAGTAT 240

QY 241 CAACATCTGTGATTTAACTTGGCGGATGGTGTAAATATGCTACTTGTCAATTTATTTCT 300
 Db 241 CAACAGCTCAATTTAGTTAGTGTGCTGCTACTTATGCTACTTGTGTTAAATTTGAATTTCT 300

QY 301 GGTGAAGAAATTCACAAATTTTCTACATTAACATGCTACTGTGAACGAGCGTTTGAATCA 360
 Db 301 GCCAAGAGTTTACTACTTTTCTAGTGTGCTACTGTACTGTACTGTACTGTACTGTACTGCT 360

QY 361 TCCATTAAGGCAATTTGGTACAGTTACTTTTACCAATTTGCAATTTGTTGGTGGACAGGT 420
 Db 361 GACACCAAGCCATAGGAACGTAAACATTTCTTCTCAATCAGTGTGGGGGATCAGGT 420

QY 421 TCATCAACTGATTTGGAAGATTTCTAAATGTTTTTACTGCTGGTACCAATACAGTCAATTT 480
 Db 421 TCAGATGTTGATTTGGCAATTTCTCAATGTTTACTGTCAGGAATCAATACAGTACTTTT 480

QY 481 AATGATGCTGATAAAGATATCTCAATTTGATGTTGAGTTTGAAGTCAACCGTTGATCCA 540
 Db 481 AATGATGCTGACACTAGCATTTCCACAACAGTTGATTTTGAAGAAATCAACCGTGCCTCC 540

QY 541 AGTGCATATTTGTTATGCTTCCAGAGTTATGCAAGTCTCAATAAGAGTGCACACTCTTTT 600
 Db 541 AGCGATCGTATCTTCTGTTCAAGAATTTTACCACGCTCTTCCACAAGCAGTAAATCTTTT 600

QY 601 GTGGCACCACAAATGTGAAATGGTTACACATCTGTGTACAAATGGGTTCTTCCAGTAGTAAC 660
 Db 601 CTTCCCAAGAATGTGCAATGGTTATCTTCTGTGTACAAATGGGATTTTCGACTGCTGTT 660

QY 661 GGTGACGTTGCTATTGATTGCTCAAAATTTATATTTATTTGTTATGTTATGATTAATGAT 720
 Db 661 ACTGGTGTCTACTATAGATTGTTCCACAGTTTCATGTCCGGATATCAATGGGTTGAATGAT 720

QY 721 TGGAAATATCCGGTTTTCATCTGAATCATTTAGTTACACTAAACATTTGTACATCTAATGA 780
 Db 721 TGGAAATATCCAAATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780

QY 781 ATTCAGATTAAATATCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
 Db 781 GTTTAGTAACTTTTCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840

QY 841 TCTGTACAGATTGTTAAACAATATATCTTTAGCATATACCAATGATTTATCTTTGCTGCTGC 900
 Db 841 TCTGCAACACAGTGCAGCTCATATACCATGCAATACACATATATATATATATATGCTTTG 900

QY 901 AGTCTGTCGAAAGTAAACCTTTCACTTTAAGTAGTGGATGATACAAAGATAGTAGTGCC 960
 Db 901 GCGGCTTCTGTTGATGACTCATTTTACTTCTGCGGGGATATAGTAATAGTAGCAAGCT 960

QY 961 GGATCTAAGGTTATGTTGCTGCTACAACTAGCAAGTGTACAGACAGTAGTACCAGTCT 1020
 Db 961 GGTCTAATGTTATGTTACCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020

QY 1021 GTCACTACTTTACCAATTCATTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1080
 Db 1021 GTGACTACTTTTACCAATTCATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080

QY 1081 CCTATTCCAACGCTACCATTCACACTTCATATGTTGGTGTGCTGCTGCTGCTGCTGCTGCT 1140
 Db 1081 CCTATTCCAACAACTACCATTCACACTTCATATGTTGGTGTGCTGCTGCTGCTGCTGCTGCT 1140

QY 1141 AAGCTGCACCAATTTGGTGAACAGCTGCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200
 Db 1141 AAACTGCACCAATTTGGTGAACAGCTGCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200

QY 1201 ACAACTGTTTACAGTGAATGGACAGGAAACATCTACACACCAACATCTGTTACCAATCCA 1260
 Db 1201 ACAACTGTTTACAGTGAATGGACAGGAAACATCTACACACCAACATCTGTTACCAATCCA 1260

QY 1261 ACTGATCAATTTGACACAGTGGTGTGACAAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320
 Db 1261 ACTGATCTATAGATAGTGTGCTGTTTCAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320

QY 1321 ACTGAATATTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 Db 1321 ACCGAGTATTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380

QY 1381 ACCGATCTGCTGATTTATCAGAGAGCC 1406
 Db 1381 ACTGACTAGTAAATTTATCAGAGAGCC 1406

RESULT 14
 CANALS42S1
 LOCUS 1407 bp DNA linear PLN 16-OCT-1998
 DEFINITION Candida albicans agglutinin-like protein (ALS4) gene, 5' partial cds.
 ACCESSION AF024586
 VERSION AF024586.1 GI:3598682
 KEYWORDS

```

SEGMENT 1 of 2
SOURCE Candida albicans.
ORGANISM Candida albicans.
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS Hoyer,L.L., Payne,T.L. and Hecht,J.E.
TITLE Identification of Candida albicans ALS2 and ALS4 and localization
of als proteins to the fungal cell surface
J. Bacteriol. 180 (20), 5334-5343 (1998)
JOURNAL 98440424
MEDLINE 9765564
PUBMED
REFERENCE 2 (bases 1 to 1407)
AUTHORS Hoyer,L.L.
TITLE Direct Submission
JOURNAL Submitted (11-SEP-1997) Veterinary Pathobiology, University of
Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES
source
Location/Qualifiers
1..1407
/organism="Candida albicans"
/strain="1161"
/db_xref="taxon:5476"
/chromosome="6"
/map="6c"
1..>1407
/gene="ALS4"
/feature="agglutinin-like sequence"
/allele="ALS4-2"
1..>1407
/gene="ALS4"
/transl_table=1
/codon_start=1
/product="agglutinin-like protein"
/protein_id="AAC64241.1"
/db_xref="GI:3598685"
/translation="MLQLLSLCVSATKVIHGVNFSNLSLFWANASYPYRGPA
TPTWTAIVGWSLDGATASAGDTFTLDMPCVFKFIDQTSIDLVDGRYATCNLSAE
EFTTSSVSCVTWTMTADTKAMGTVPFPFSGVSGSDVDLANSQCTAGINVTFT
NDGDFSISTTVDFFSTVASSDRILLRSLPSQASVLEFLPQECANGYSTGMGFS
AGTGATIDCSNVHFGISNGLANDWNPYISSEFSYTKTCTSVLTVYONVAPGVFV
DAYSATRVSSYTMQYNIYACGAASVDDSFTHWRGYSNSQASNGNITIVVTRV
TDSATVTLFNSDITDKTKIEILOPIITTTTTSYGVVTSYSTKAPIGETATVI
VDVPTVTTTTSVSEWTGTTTTTTRTNTDTSIDTVVQVPSNPVTVTTEYSQSYAT
TTVTAPPGDSDSVIIREP"
1300..1407
repeat_region
/feature="first of multiple copies of 108bp imperfect repeat"
BASE COUNT 389 a 297 c 257 g 464 t
ORIGIN
Query Match 22.1%; Score 836.4; DB 8; Length 1407;
Best Local Similarity 74.7%; Pred. No. 4e-141;
Matches 1050; Conservative 0; Mismatches 356; Indels 0; Gaps 0;
QY 1 ATGCTTCAACAATTTACATTTTATTCCTATATTTGTCATATGCAATGCAAGTGCAAGACAATC 60
DB 1 ATGCTTCAACAATTTTGTGTAAGCCCTCTGTGTATCATGCTACGCTACGCAAAAAGTATT 60
QY 61 ACTGGTGTGTTTGTAGTATTTTAAATTCATTAATTCCTGCTCAATGCTGCTAAATATGCTTTC 120
DB 61 ACAGGTGTTTTCAATAGTATTTTAAATTCGTTAACTTGGGCAATGCTGCTTCTTATCCATAT 120
QY 121 AAAGGCCAGGATACCCAACTTGGGAATGCTGTTTGGGTGCTTGTAGATGCTGCTACCACT 180
DB 121 AGAGTCCAGCTACTCCTACTTGGACCGCTGTAAATAGATGCTTCTTATAGATGGAGCTACT 180
QY 181 GCCAATCCAGGGGATACATTCATTCATATGATATGCAATGCTGCTGCTTAAATATATCTTCA 240
DB 181 GCTAGTGTGCTGGTGACACATTCACGTAGACATGCTGCTGCTGCTTAAATATATCTGAT 240
QY 241 CAACATCTGTTGATTTAACTGCCGATGGTGTAAATATGCTTAAATATCTTAAATTTATCT 300
DB 241 CAACGTCATTTGATTTAGTTAGTGTGCTGCTGCTTAAATATGCTTAAATTTGAACTCT 300

```

```

QY 301 GGTGAAGAATTCACAACATTTTCTTACATTAACATGCTACTGTGACGACGCTTTGAAATCA 360
DB 301 GCCGAAGAGTTTACTACTTTTCTTAGTGTGCTGCTACTGTGCTACTACTACTACTACTACT 360
QY 361 TCCATTAAAGGCAATTTGGTACAGTTACTTTACCAATTTGCAATTTCAATTTGGTGGGAACAGT 420
DB 361 GACACCAAGCCATCGGAACGTGAACATTTACCTTTCTCATTTCAAGTGTGGGGGAGCATGCT 420
QY 421 TCATCAACTGATTTGGAAGATTTCTAAATTTGTTTACTGCTGGTGGTACCAATACATGCTCAT 480
DB 421 TCAGATGTTGATTTGGCAAAATTTCTCAATTTGTTTACTGCGGGAATCAATACAGTTACTTT 480
QY 481 AATGATGCTGATAAAGATATCTCAATTTGATTTGAGTTTGAAGTTTGAAGTTTGAAGTTTGA 540
DB 481 AATGATGCTGACATAGCATTTCCACACAGTTGATTTGAAATTTGAAGTTTGAAGTTTGA 540
QY 541 AGTGCAATTTTGTATGCTTCCAGATTTATGCAAGTCTCAATTAAGGTCAACACTCTTTT 600
DB 541 AGCGATGCTATCTTCTGTCGAAGATTTTACCAGCTCTTTCACAAGCAGTAAGTCTTTT 600
QY 601 GTGGCACCACATGTAAGATTTTACACATCTGCTACAAATGGGTTTCTCCAGTAGTAAC 660
DB 601 CTTCCCAAGATGTCGAATTTTATCTTCTGCTACAAATGGGATTTTCAACTGCTGCT 660
QY 661 GGTGACGTTGCTATTTGCTCAATATTTCAATTTGATTTGCTATTCACAAAGGATTAATGAT 720
DB 661 ACTGCTGCTATATAGATTTTCCAGCTTCAATGCTGCGGATATCAATTTGGGTTGATGAT 720
QY 721 TGGAAATATCCGGTTTTCATCTGAATTTAGTTAGTACACTTAACTTTGATACATCTAATGGA 780
DB 721 TGGAAATATCCAAATTTTCACTGGAATTTCTTTTACACAAGACCTGTATACATCAACAAGT 780
QY 781 ATTCAGATTAATATCAAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 781 GTTTTAGTAACATTTATCAGAATTTCTGCTCGGATATCGTCCATTTGCTGATGCTTATGTT 840
QY 841 TCTGCTACAGATTTTAAACAATATATCTTTAGCATATACCAATGATTTATTTGCTGCTGCTG 900
DB 841 TCTGCAACAGCAGTCACTCATATACCATGCAATACACTATATATATGCTTGTGCTGCTG 900
QY 901 AGTCTGTGCAAGTAAACCTTTTCACTTTAGATGGAGTGGATACAGAATAGTATGATGCC 960
DB 901 GCGGCTTCTGTTGATGACTCATTTTACTTACTTGGCGGGATATAGTAATAGTCAAGCT 960
QY 961 GGATCTAAGGTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 961 GGTTCCTAATGTTTACCATTGTTGTAACAATGTTGTTGTAACAATGTTGTTGTAACAATGTT 1020
QY 1021 GTCATCTTTTACCATTGTTGTAACAATGTTGTTGTAACAATGTTGTTGTAACAATGTTGTA 1080
DB 1021 GTGACTCTTTTACCATTGTTGTTGTAACAATGTTGTTGTAACAATGTTGTTGTAACAAT 1080
QY 1081 CCTATTCCAACTTACCATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1140
DB 1081 CCTATTCCAACTTACCATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1140
QY 1141 AAGACTGCAACCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200
DB 1141 AAGACTGCAACCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1200
QY 1201 AAGACTGTTTACCAGTGAATGGAGCAATTCATACCAACCACTGCTGCTGCTGCTGCTGCTGCT 1260
DB 1201 AAGACTGTTTACCAGTGAATGGAGCAATTCATACCAACCACTGCTGCTGCTGCTGCTGCTGCT 1260
QY 1261 ACTGATTTCAATTTGACACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1320
DB 1261 ACTGATTTCAATTTGACACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1320
QY 1321 ACTGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1321 ACCGAGTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380

```

181	GCTGACCCAGGAGACATTCACCTTGATTTGGTCCCTGGTGTATTTAAATTTATAACTACC	240
241	CAACACATCTGTTGATTTAACTGCCGATGGTTAAATATAGCTACTTGTCAATTTTATCT	300
241	CAACACATCTGTTGATTTGACTGCTGATGCTGTAGCTATGCCACTTGTGACTTTAATGCT	300
301	GGTGAAGAAATTCACAACCTTTCTACATTTAAGATGACTGTGAACGACGCTTTGAAATCA	360
301	GGTGAAGAAATTTACGACATTTTCTCCCTTATCATGTACTGTGAACAGTGTTCGGTATCA	360
361	TCCATTAAAGGCATTTGGTACAGTACTTTTACCAATTTGCAATTCAAATCTGGTGGAAACAGG	420
361	TATGCTAGGTTTCTGGTAGCGTCAAAATTCGCCATTTACATTTCAATTTAGTGGGAACAGG	420
421	TCATCAACTGATTTGGAAGATTTCTAAATGTTTTACTGCTGGTACCAATACAGTFCACATTT	480
421	TCCTCAGTTGATTTGGCAGATTTCCAAATGTTTTACTGCCGGAACAAACACTGTGACTTTC	480
481	AATGATGGTGATAAAGATATCAATTGATGTTGAGTTTGAAGTCAACCGTTGATGCCA	540
481	ATGAGATGGCGATACAAAGATTTCTACCACTGTTGATTTTGACGCGTCTCCAGTATCACCC	540
541	AGTGCAATTTCTATGCTTCCAGAGTTATGCCAAGTCTCAATAAAGTFCACAACCTCTTTT	600
541	AGTGGTTATTTACAAAGCTACGAATTTATCTCTAGTCTCAATAAACTTTCAAGCTTTTC	600
601	GTGGCACCACAAATGTGAAATGGTTACACATCTGGTACAAATGGGTTCTCCAGTAGTAAC	660
601	GTGGTGCCACAAATGTGAGAAATGGTTACACATCTGGTATATAGGGATTTGTAGCTAGTAAC	660
661	GGTGACGTTGCTATTGATGTCTCAAAATATTCATATGGTATTCACAAAAGATTTAAATGAT	720
661	GCT---GCTACTATTGATGGCTCAAAATGTCAATATAGGAATATCAAAAGGTTTAAATGAT	717
721	TGGAATTTATCCGGTTTCTCTCAATCATTTAGTTACACTTAAACCTTGTACATCTAATGGA	780
718	TGGAATTTTCCAGTAAAGTTCGAATCATTTTCTTACACGAAACTTTGTACGTCAACCCAGT	777
781	ATTACAGATTTAAATATCAAAATGTACCTGCTGGTTATCTGCTCCATTTATGATGCTTATAT	840
778	ATTACAGTTGAATTTCAAAATGTCTGCTGGGTATCGCCCTTTTGGTTGATGCAATATTT	837
841	TCTGCTACAGATTTAAACCAATATCTTTTAGCATATACCAAATGATTATCTTGTGCTGGC	900
838	TCTGCGAAGAAATATTGATAATATACCTTGACGTACGCAAAATGAGTATACTTGTGAAAT	897
901	AGTCGCTGCAAAAGTAAACCTTTCACCTTTAAGATGGACTGGGATACAGATATAGTGCC	960
898	GGCAATACTGTGGTTGATCCATTTACTTTTAACTGGTGGGGTATAAAACTCTGAAGCA	957
961	GGATCTAACGGTATTGTCTATTGTGTACAACCTTAGAAACAGTTACACACAGTACCACTGCT	1020
958	GACTCTGACGGGGATGTGATCGTAGTTACAAACCGAATCTGCACAGACAGTACACAGCT	1017
1021	GTCACTACTTTACCATTCAATCCAAGTGTGTGATAAACCAAAACAAATCGAAATTTTGCAA	1080
1018	GTGACTACTTTACCTTTCAATCCAAGTGTGATATAAACCGAAGTATAAAACTCTTGCAA	1077
1081	CCTATTCCAAACCACTACCATCAACTTCAATATGTTGGTGTGACTACTTCCCTATCTGACT	1140
1078	CCCATTTCCACGACCAAAATTAACACTTCATATTTGGTATTTTCCACTTCCCTATGAANCA	1137
1141	AAAGCTGCACCAATTTGGTCAACAGCTACTGTTATTGTTGATGTGCGCATATCATACTACC	1200
1138	TTAAACCGGAACAATTTGGTGGTACTGCGCAGCTCATTTGCGTACACCTTATCATATCACT	1197
1201	ACAACCTGTTACCAGTGAATGGACGAGGAACAATCACTACCACCAAACTCGTACCACTCCA	1260
1198	GCCACTGTTACAANAATTCGGACTGGGTCAATTTACAACCTACCAGTACTTATATACTAATCC	1257
1261	ACTGATTCATTAATGACACAGTGGTGGTACAGTTTCCACTGCGCAAAATCCAAGTGTGTACT	1320
1258	ACTGGTTCCATAGACACTGTTATTTGTGCAAAATTCCACTGCTGATGCTCAACTCAACTATA	1317

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 01:57:43 ; Search time 530 Seconds
(without alignments)
16086.916 Million cell updates/sec

Title: US-09-715-876-7

Perfect score: 3786

Sequence: 1 atgtctcaacaattacatt.....tgctctgttcatttagtga 3786

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	359.2	9.5	424	17 AAT29067
2	108	2.9	108	17 AAT32337
3	108	2.9	108	17 AAT32330
4	108	2.9	108	17 AAT32331
5	108	2.9	108	17 AAT32332
6	108	2.9	108	17 AAT32334
7	108	2.9	108	17 AAT32336
8	108	2.9	108	17 AAT29063
9	106.4	2.8	108	17 AAT32333

10	106.4	2.8	108	17 AAT32335	Probe for Candida
11	99	2.6	99	17 AAT32338	Probe for Candida
c 12	94.2	2.5	8201	21 AAA88864	Human dentin sialo
c 13	94.2	2.5	8201	24 ABQ73537	Human dentin sialo
14	87	2.3	5511	21 AAA61847	Cryptosporidium pa
15	87	2.3	5511	24 ABT04776	C-parvum Gp900 gen
16	87	2.3	7334	21 AAA61846	Cryptosporidium pa
17	87	2.3	7334	24 ABT04775	C parvum Gp900 gen
18	86.2	2.3	5163	19 AAV20700	Cryptosporidium pa
19	86.2	2.3	5163	21 AAA61849	ORF encoding a por
20	86.2	2.3	5163	24 ABT04778	C parvum Gp900 gen
21	86.2	2.3	5318	19 AAV20701	Cryptosporidium pa
22	86.2	2.3	5318	21 AAA61848	DNA encoding a por
23	86.2	2.3	5318	24 ABT04777	C parvum Gp900 gen
24	85.2	2.3	108	17 AAT29064	Probe for Candida
c 25	82.4	2.2	100	15 AAQ62588	Candida albicans-s
c 26	82.4	2.2	100	15 AAQ62590	Candida albicans-s
c 27	81.4	2.2	7758	24 ABL33102	Human immune syste
c 28	76	2.0	100	15 AAQ62589	Candida albicans-s
c 29	74.4	2.0	267156	24 ABL68560	Kidney cancer rela
c 30	68.4	1.8	40875	18 AAT80043	Insert from cosmid
c 31	68	1.8	100	15 AAQ62592	Candida albicans-s
c 32	68	1.8	7758	24 ABL33103	Human immune syste
c 33	67.6	1.8	2336	23 ABL25662	Drosophila melanog
c 34	64.8	1.7	2790	7 AAN60473	Sequence encoding
c 35	62.6	1.7	1368	10 AAN91235	DNA sequence of pr
c 36	62.2	1.6	2014	24 ABA90791	Bacillus anthracis
c 37	58	1.5	3168	20 AAX36552	C. albicans Rbt1 c
c 38	58	1.5	3983	23 ABL15838	Drosophila melanog
c 39	57.4	1.5	2535	23 ABL25476	Drosophila melanog
c 40	57.2	1.5	100	15 AAQ62593	Candida albicans-s
c 41	57.2	1.5	3399	17 AAT05868	Chicken leucocytoz
c 42	57.2	1.5	32392	24 ABL56203	Ampev genome fragm
c 43	57.2	1.5	50000	24 ABL55643	Ampev genome fragm
c 44	57	1.5	100	15 AAQ62591	Candida albicans-s
c 45	56.6	1.5	14987	24 ABL32630	Human immune syste

ALIGNMENTS

RESULT 1

AAT29067

ID AAT29067 standard; DNA; 424 BP.

XX

AC AAT29067;

XX

DT 28-NOV-1996 (first entry)

XX

DE Probe for Candida albicans and Candida stellatoideis.

XX

KW Probe: primer; detection; identification; Candida albicans;

KW Candida stellatoideis; sputum; bronchial washings; blood; milk;

KW lymph fluid; skin; soft tissue; ss.

XX

OS Synthetic.

XX

PN WO9618745-Al.

XX

PD 20-JUN-1996.

XX

PF 08-DEC-1995; 95WO-US16153.

XX

PR 16-DEC-1994; 94US-0357962.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Hoyer LL, Livi GP, Shatzman A;

XX WPI; 1996-300661/30.

XX

PT C. albicans and C. stellatoideis specific probes and primers - for

PT specific detection of Candida infection

CC sequences (AAT29065, AAT29066) as amplification primers and then
CC contacting one of the probe sequences with the amplified product and
CC detecting hybridisation.

XX Sequence 108 BP; 31 A; 30 C; 15 G; 32 T; 0 other;

Query Match 2.9%; Score 108; DB 17; Length 108;

Best Local Similarity 100.0%; Pred. No. 2.4e-15;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1408 CCAACCATAGTCTCACTACTACTGATATATGGTCACAAATCCTTTGCTACTACTACT 1467

DB 1 CCAACCATAGTCTCACTACTACTGATATATGGTCACAAATCCTTTGCTACTACTACT 60

OY 1468 GTTACTGCTCTCCAGGTGGTACTGACTCAGTCAATTAATATCAGAGAACCA 1515

DB 61 GTTACTGCTCTCCAGGTGGTACTGACTCAGTCAATTAATATCAGAGAACCA 108

RESULT 4

AAT32331

ID AAT32331 standard; DNA; 108 BP.

XX

AC AAT32331;

XX

DT 29-NOV-1996 (first entry)

XX

DE Probe for Candida albicans and Candida stellatoideis.

XX

KW Probe; primer; detection; identification; Candida albicans;

KW Candida stellatoideis; sputum; bronchial washings; blood; milk;

KW lymph fluid; skin; soft tissue; ss.

XX

OS Synthetic.

XX

PN WO9618745-A1.

XX

PD 20-JUN-1996.

XX

PF 08-DEC-1995; 95WO-US16153.

XX

PR 16-DEC-1994; 94US-0357962.

XX

PA (SMIK) SMITHKLINE BEECHAM CORP.

XX

PI Hoyer LL, Livi GP, Shatzman A;

XX

DR WPI; 1996-300661/30.

XX

PT C. albicans and C. stellatoideis specific probes and primers - for

PT specific detection of Candida infection

XX

PS Claim 1; Figure 5A; 33pp; English.

XX

CC Five synthetic sequences (AAT29063-67) are used as probes to detect

CC the presence of Candida albicans and Candida stellatoideis in

CC biological fluids e.g. sputum, bronchial washings, blood, milk and

CC lymph fluid or in tissue samples e.g. skin and soft tissues. The

CC method of detection may also comprise using two of the probe

CC sequences (AAT29065, AAT29066) as amplification primers and then

CC contacting one of the probe sequences with the amplified product and

CC detecting hybridisation.

XX

Sequence 108 BP; 30 A; 31 C; 16 G; 31 T; 0 other;

Query Match

Best Local Similarity 2.9%; Score 108; DB 17; Length 108;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1516 CCAATCACTGTCACTACACCGAGTATGGTCTCAATCCTTTGCTACTACTACA 1575

DB 1 CCAATCACTGTCACTACACCGAGTATGGTCTCAATCCTTTGCTACTACTACA 60

OY 1576 GTTACTGCTCTCCAGGTGGTACTGACTCAGTCAATTAATATCAGAGAACCT 1623

DB 61 GTTACTGCTCTCCAGGTGGTACTGACTCAGTCAATTAATATCAGAGAACCT 108

RESULT 5

AAT32332

ID AAT32332 standard; DNA; 108 BP.

XX

AC AAT32332;

XX

DT 29-NOV-1996 (first entry)

XX

DE Probe for Candida albicans and Candida stellatoideis.

XX

KW Probe; primer; detection; identification; Candida albicans;

KW Candida stellatoideis; sputum; bronchial washings; blood; milk;

KW lymph fluid; skin; soft tissue; ss.

XX

OS Synthetic.

XX

PN WO9618745-A1.

XX

PD 20-JUN-1996.

XX

PF 08-DEC-1995; 95WO-US16153.

XX

PR 16-DEC-1994; 94US-0357962.

XX

PA (SMIK) SMITHKLINE BEECHAM CORP.

XX

PI Hoyer LL, Livi GP, Shatzman A;

XX

DR WPI; 1996-300661/30.

XX

PT C. albicans and C. stellatoideis specific probes and primers - for

PT specific detection of Candida infection

XX

PS Claim 1; Figure 5A; 33pp; English.

XX

CC Five synthetic sequences (AAT29063-67) are used as probes to detect

CC the presence of Candida albicans and Candida stellatoideis in

CC biological fluids e.g. sputum, bronchial washings, blood, milk and

CC lymph fluid or in tissue samples e.g. skin and soft tissues. The

CC method of detection may also comprise using two of the probe

CC sequences (AAT29065, AAT29066) as amplification primers and then

CC contacting one of the probe sequences with the amplified product and

CC detecting hybridisation.

XX

Sequence 108 BP; 34 A; 36 C; 16 G; 22 T; 0 other;

Query Match

Best Local Similarity 2.9%; Score 108; DB 17; Length 108;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1624 CCAACCACTGTCACTACCACTCAATATTTGGTCCCAATCTTACGCAACCACTACT 1683

DB 1 CCAACCACTGTCACTACCACTCAATATTTGGTCCCAATCTTACGCAACCACTACT 60

OY 1684 GTGACTGCTCTCCAGGAGGCACTGACTCAGTCAATTAATATCAGAGAACCA 1731

DB 61 GTGACTGCTCTCCAGGAGGCACTGACTCAGTCAATTAATATCAGAGAACCA 108

RESULT 6

AAT32334

ID AAT32334 standard; DNA; 108 BP.

XX

AC AAT32334;

XX

DT 29-NOV-1996 (first entry)

XX

DE Probe for Candida albicans and Candida stellatoideis.

CC biological fluids e.g. sputum, bronchial washings, blood, milk and
 CC lymph fluid or in tissue samples e.g. skin and soft tissues. The
 CC method of detection may also comprise using two of the probe
 CC sequences (AAT29065, AAT29066) as amplification primers and then
 CC contacting one of the probe sequences with the amplified product and
 CC detecting hybridisation.

XX Sequence 108 BP; 29 A; 29 C; 19 G; 31 T; 0 other;

Query Match 2.9%; Score 108; DB 17; Length 108;
 Best Local Similarity 100.0%; Pred. No. 2.4e-15;
 Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1300 CCAATCCCAACTGTTAGTACTACTGAATATTGCTCTCAGTCTTTGCTACAAACCACCTACA 1359

Db 1 CCAATCCCAACTGTTAGTACTACTGAATATTGCTCTCAGTCTTTGCTACAAACCACCTACA 60

QY 1360 GTTACTGCTCCCTCCAGGTGGTACCGATATGATATCAGAGAGCCA 1407

Db 61 GTTACTGCTCCCTCCAGGTGGTACCGATATGATATCAGAGAGCCA 108

RESULT 9

AAT32333

ID AAT32333 standard; DNA; 108 BP.

XX AC AAT32333;

DT 29-NOV-1996 (first entry)

XX Probe for Candida albicans and Candida stellatoideis.

XX Probe: primer; detection; identification; Candida albicans;

KW Candida stellatoideis; sputum; bronchial washings; blood; milk;

KW lymph fluid; skin; soft tissue; ss.

XX Synthetic.

XX WO9618745-A1.

XX 20-JUN-1996.

XX 08-DEC-1995; 95WO-US16153.

XX 16-DEC-1994; 94US-0357962.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Hoyer LL, Livi GP, Shatzman A;

XX WPI; 1996-300661/30.

XX C. albicans and C. stellatoideis specific probes and primers - for
 PT specific detection of Candida infection

XX Claim 1; Figure 5A; 33pp; English.

XX Five synthetic sequences (AAT29063-67) are used as probes to detect
 CC the presence of Candida albicans and Candida stellatoideis in
 CC biological fluids e.g. sputum, bronchial washings, blood, milk and
 CC lymph fluid or in tissue samples e.g. skin and soft tissues. The
 CC method of detection may also comprise using two of the probe
 CC sequences (AAT29065, AAT29066) as amplification primers and then
 CC contacting one of the probe sequences with the amplified product and
 CC detecting hybridisation.

XX Sequence 108 BP; 34 A; 33 C; 16 G; 25 T; 0 other;

Query Match

Best Local Similarity 2.8%; Score 106.4; DB 17; Length 108;
 Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1732 CCAAACCCACACTGTCTACTACTGAATACTGTGTCACATCATATGCCACCACCTACTACC 1791

Db 1 CCAAACCCACACTGTCTACTACTGAATACTGTGTCACATCATATGCCACCACCTACTACC 60

QY 1792 GTAACCTGCACCCAGGTGGTACTGACACTGTATCATATAGAGAGCCA 1839

Db 61 GTAACCTGCACCCAGGTGGTACTGACACTGTATCATATAGAGAGCCA 108

RESULT 10

AAT32335

ID AAT32335 standard; DNA; 108 BP.

XX AC AAT32335;

DT 29-NOV-1996 (first entry)

XX Probe for Candida albicans and Candida stellatoideis.

XX Probe: primer; detection; identification; Candida albicans;

KW Candida stellatoideis; sputum; bronchial washings; blood; milk;

KW lymph fluid; skin; soft tissue; ss.

XX Synthetic.

XX WO9618745-A1.

XX 20-JUN-1996.

XX 08-DEC-1995; 95WO-US16153.

XX 16-DEC-1994; 94US-0357962.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Hoyer LL, Livi GP, Shatzman A;

XX WPI; 1996-300661/30.

XX C. albicans and C. stellatoideis specific probes and primers - for
 PT specific detection of Candida infection

XX Claim 1; Figure 5A; 33pp; English.

XX Five synthetic sequences (AAT29063-67) are used as probes to detect
 CC the presence of Candida albicans and Candida stellatoideis in
 CC biological fluids e.g. sputum, bronchial washings, blood, milk and
 CC lymph fluid or in tissue samples e.g. skin and soft tissues. The
 CC method of detection may also comprise using two of the probe
 CC sequences (AAT29065, AAT29066) as amplification primers and then
 CC contacting one of the probe sequences with the amplified product and
 CC detecting hybridisation.

XX Sequence 108 BP; 33 A; 37 C; 14 G; 24 T; 0 other;

Query Match

Best Local Similarity 2.8%; Score 106.4; DB 17; Length 108;
 Matches 107; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1948 CCAAACCCACACTGTCTACCACTACTGAATACTGTGTCATCATATGCCACCACCTACTACC 2007

Db 1 CCAAACCCACACTGTCTACCACTACTGAATACTGTGTCATCATATGCCACCACCTACTACC 60

QY 2008 ATTACCGCTCCCACTGCTGAACCTGATACCGTTCTTATCAGAGAGCCA 2055

Db 61 ATTACCGCTCCCACTGCTGAACCTGATACCGTTCTTATCAGAGAGCCA 108

RESULT 11

AAT32338

ID AAT32338 standard; DNA; 99 BP.

XX AC AAT32338;

DT 29-NOV-1996 (first entry)
 DE Probe for Candida albicans and Candida stellatoideis.
 XX
 KW Probe: primer; detection; identification; Candida albicans;
 KW Candida stellatoideis; sputum; bronchial washings; blood; milk;
 KW lymph fluid; skin; soft tissue; ss.
 XX
 OS Synthetic.
 OS
 PN WO9618745-A1.
 PN
 XX 20-JUN-1996.
 XX
 XX 08-DEC-1995; 95WO-0516153.
 XX
 XX 16-DEC-1994; 94US-0357962.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 XX Hoyer LL, Livi GP, Shatzman A;
 XX
 XX WPI; 1996-300661/30.
 XX
 XX C. albicans and C. stellatoideis specific probes and primers - for
 PT specific detection of Candida infection
 PT
 XX Claim 1; Figure 5A; 33pp; English.
 XX
 CC Five synthetic sequences (AAT29063-67) are used as probes to detect
 CC the presence of Candida albicans and Candida stellatoideis in
 CC biological fluids e.g. sputum, bronchial washings, blood, milk and
 CC lymph fluid or in tissue samples e.g. skin and soft tissues. The
 CC method of detection may also comprise using two of the probe
 CC sequences (AAT29065, AAT29066) as amplification primers and then
 CC contacting one of the probe sequences with the amplified product and
 CC detecting hybridisation.
 XX
 XX Sequence 99 BP; 29 A; 29 C; 14 G; 27 T; 0 other;
 SQ
 Query Match 2.6%; Score 99; DB 17; Length 99;
 Best Local Similarity 100.0%; Pred. No. 2.7e-13;
 Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2272 CCAATCCACAGTTACTACTACTGTAATATTTGTCACATCATTTGCCACACCCACCA 2331
 Db 1 CCAATCCACAGTTACTACTACTGTAATATTTGTCACATCATTTGCCACACCCACCA 60
 QY 2332 GTTACTGTCCTCCAGGTGGTACTGACACTGTGATTATC 2370
 Db 61 GTTACTGTCCTCCAGGTGGTACTGACACTGTGATTATC 99
 RESULT 12
 AAA88864/c
 ID AAA88864 standard; DNA; 8201 BP.
 XX
 AC AAA88864;
 XX
 DT 19-FEB-2001 (first entry)
 XX
 DE Human dentin sialophosphoprotein gene.
 XX
 KW Dentin sialophosphoprotein; DSPP; human; SIBLINGS;
 KW Integrin-binding ligand; diagnosis; antiinflammatory; therapy;
 KW osteoporosis; chromosome 4; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 2387..7896
 FT CDS /tag= a
 FT /note= "contains exons"

FT exon 2387..2437
 FT /tag= b
 FT Intron 2438..3576
 FT /tag= c
 FT exon 3577..3660
 FT /tag= d
 FT Intron 3661..3793
 FT /tag= e
 FT exon 3794..4780
 FT /tag= f
 FT Intron 4781..5256
 FT /tag= g
 FT exon 5257..7896
 FT /tag= h
 XX
 PN WO200062065-A1.
 XX
 XX 19-OCT-2000.
 XX
 XX 07-APR-2000; 2000WO-US09349.
 XX
 XX 09-APR-1999; 99US-0128468.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA
 XX Fisher LW, Fedarko NS, Young MF;
 XX
 XX WPI; 2000-679515/66.
 XX P-PSDB; AAB19772.
 DR
 XX
 XX Detecting small integrin-binding ligand N-linked glycoproteins for
 PT detection of a tumor or protection against a complement mediated immune
 PT response, comprises detection where Factor H is not an inhibitor -
 XX
 PS Disclosure; Page 98-100; 110pp; English.
 XX
 CC The present sequence is that of DNA encoding human dentin
 CC sialophosphoprotein (DSPP, see AAB19772), a member of the small
 CC integrin binding ligand, N-linked glycoproteins (SIBLINGS) family.
 CC The invention provides methods and compositions for exploiting the
 CC discovery that members of the SIBLINGS family bind to complement
 CC Factor H, conferring resistance to complement mediated lysis. A
 CC claimed method of conferring protection against a complement
 CC mediated immune response involves providing a reservoir or other
 CC supply in the subject's body so that a SIBLINGS protein can be
 CC dispersed to interfere with complement mediated lysis and
 CC inflammation. This protects cells that are grafted onto foreign
 CC tissue or bone marrow cells introduced into a foreign host. The
 CC SIBLINGS protein can be BSP, OPN, DMP1 or DSPP. A method of
 CC detecting a SIBLINGS protein in a sample from a subject suspected
 CC of having abnormal bone turnover, especially osteoporosis, is also
 CC claimed.
 XX
 SQ Sequence 8201 BP; 2801 A; 1477 C; 1873 G; 2050 T; 0 other;
 Query Match 2.5%; Score 94.2; DB 21; Length 8201;
 Best Local Similarity 41.7%; Pred. No. 1.6e-11;
 Matches 598; Conservative 0; Mismatches 823; Indels 0; Gaps 0;
 QY 952 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGACACATTACAGACGT 1011
 Db 7290 ACTGCTGTCACCTGCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7231
 QY 1012 ACCACTGCTGCTCAGTACTTACCAATCAATCAAGTGTGATAAAACCAACAAATCGAA 1071
 Db 7230 ATCGCTGCTGCTCAGTACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7171
 QY 1072 ATTTTGCAACCTATTCCAACTACCACTACCAATCACAACCTTATATGTTGGTGACTACTTCC 1131
 Db 7170 GCTGTCACCTGCTATCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTATC 7111
 QY 1132 TATCTGACTAAGACCTGACCAATTTGGTGAACAGCTACTGTTATGTTGATGTCATAT 1191

```
Db 7110 ACTGCTGCTACTGCTATCGCTGCTGTTGCTGCTGCTACTGCTGCTGCTGCTATTTGCTATC 7051
QY 1192 CATACTACCAACTGTTTACCAGTGAATGGAGCAAGAACTACTACCAACCACTACCTGCT 1251
Db 7050 ACTGCTGCTACTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6991
QY 1252 ACCAATCCAACTGATTCAATTTGACACAGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1311
Db 6990 GTCAGTGTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6931
QY 1312 GTTAGTACTAGTGAATATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1371
Db 6930 ACTATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6871
QY 1372 CCAGTGTGATCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1431
Db 6870 GCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6811
QY 1432 GAATATTGGTCAAAATCCCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1491
Db 6810 GCTGCTTTCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6751
QY 1492 GACTCAGTAAATATCAGAGAACCAACCAATCCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1551
Db 6750 GTCAGTGTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6691
QY 1552 CAATCCCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1611
Db 6690 ATCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6631
QY 1612 ATCAGAGAACCTCCAAACCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1671
Db 6630 GCTGCTACTATTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6571
QY 1672 ACCAAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1731
Db 6570 GCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6511
QY 1732 CCAAAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1791
Db 6510 ACTGCTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6451
QY 1792 GTAAGTGCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1851
Db 6450 GTCAGTGTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6391
QY 1852 GTCAGTACTAGTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1911
Db 6390 GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6331
QY 1912 CCAAGTGGCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1971
Db 6330 ACTATCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6271
QY 1972 GAATAGTGGTCTCAATCATATATGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2031
Db 6270 GTCAGTACTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6211
QY 2032 GATACCGTCTTATCAGAGAGCCCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2091
Db 6210 GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6151
QY 2092 CAATCATATGCTTACAAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2151
Db 6150 ACTGCTGCTACTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6091
QY 2152 ATCAGAGAGCCCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2211
Db 6090 GTCAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6031
QY 2212 ACAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2271
Db 6030 GTCGCTCTCTGATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5971
```

```
QY 2272 CCAATCCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2331
Db 5970 GTCAGTGTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5911
QY 2332 GTTAGTGTCTCCAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2362
Db 5910 GTCAGTGTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5880

RESULT 13
ABQ73537/c
ID ABQ73537 standard; DNA; 8201 BP.
XX
AC ABQ73537;
DT 03-OCT-2002 (first entry)
XX
DE Human dentin sialophosphoprotein precursor (DSPP) gene SEQ ID NO:1.
XX
KW Human; dentin sialophosphoprotein precursor; dentin sialophosphoprotein;
KW DSPP; dentinogenesis imperfecta type II; deafness; auditory;
KW chromosome 4q21; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT promoter 1..6
FT exon /*tag= a
FT exon /*tag= b
FT intron /*number= 1
FT intron /*tag= c
FT exon /*number= 1
FT exon /*tag= d
FT exon /*number= 2
FT CDS /*tag= e
FT /*product= "dentin sialophosphoprotein precursor"
FT /*note= "contains introns"
FT sig_peptide 2387..2431
FT /*tag= f
FT mat_peptide 2432..7893
FT /*tag= g
FT /*product= "dentin sialophosphoprotein"
FT /*note= "contains introns"
FT mat_peptide 2432..5520
FT /*tag= h
FT /*product= "dentin sialoprotein"
FT /*note= "contains introns"
FT intron 2438..3576
FT /*tag= i
FT exon /*number= 2
FT /*tag= j
FT intron /*number= 3
FT /*tag= k
FT exon /*number= 3
FT /*tag= l
FT intron /*number= 4
FT /*tag= m
FT exon /*number= 4
FT /*tag= n
FT mat_peptide 5521..7893
FT /*tag= o
FT /*product= "phosphophorin"
```

FT misc_feature 5596..5604
 FT /tag- p
 FT /note- "Cell binding domain"
 FT polyA_signal 7988..7993
 FT /tag- g
 FT polyA_signal 8171..8176
 FT /tag- r
 XX
 PN WO200258722-A1.
 XX
 PN 01-AUG-2002.
 XX
 PF 30-AUG-2001; 2001WO-CN01292.
 XX
 PR 05-SEP-2000; 2000CN-0125042.
 XX
 XX (SHAN-) SHANGHAI RES CENT BIOTECHNOLOGY.
 PA Kong X, Xiao S, Zhao G, Yu C, Hu L;
 XX WPI; 2002-557897/59.
 DR P-PSDB; ABP51785.
 DR
 XX
 PT Diagnosis of dentinogenesis imperfecta type III and its accompanying
 PT deafness using dentin sialophosphoprotein gene and encoded products -
 XX
 PS Example 3; Page 23-27; 38pp; Chinese.
 XX
 CC The present invention describes a method (M1) for the diagnosis of
 CC dentinogenesis imperfecta type II and/or its accompanying deafness
 CC comprising determining the dentin sialophosphoprotein (DSPP) gene, its
 CC transcript and/or protein of an individual for comparison of their
 CC sequences with the normal sequences and judging the individual to have
 CC higher risk of suffering from the disease than the normal population.
 CC Also described are: (1) treating dentinogenesis imperfecta type III
 CC and/or its accompanying deafness by administering a safe and effective
 CC dose of normal DSPP and/or DSP protein to patients; (2) drug compositions
 CC containing safe doses of DSPP and/or DSP protein; and (3) a reagent kit
 CC for detecting dentinogenesis imperfecta type II and/or its accompanying
 CC deafness containing primers for specific amplification of DSPP gene or
 CC its transcript, or containing probes for binding to the mutation site.
 CC The DSPP gene and protein sequences have auditory activity. The method
 CC (M1), dentin sialophosphoprotein (DSPP) gene and DSP protein are useful
 CC for diagnosing and treating imperfecta type II and/or its accompanying
 CC deafness. The DSPP gene is located to chromosome 4q21. The present
 CC sequence represents the human DSPP gene from the present invention.
 XX
 SX Sequence 8201 BP; 2801 A; 1477 C; 1873 G; 2050 T; 0 other;
 SQ
 Query Match 2.5%; Score 94.2; DB 24; Length 8201;
 Best Local Similarity 41.7%; Pred. No. 1.6e-11;
 Matches 588; Conservative 0; Mismatches 823; Indels 0; Gaps 0;
 QY 952 AGTGATCGCGGATCTAACGGTATTGTCATTTGCTGACAACTAGAACACTTACAGACAGT 1011
 DB 7290 ACTGCTGCTACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7231
 QY 1012 ACCACTGCTGCTCACTACTTACCAATCAATCAAGTTGATAAACCAACAAACATCGAA 1071
 DB 7230 ATCGTGCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7171
 QY 1072 ATTTTGCAACCTATTCCCAACCACTACCAATCACTACCACTTCAATATTTGGTGAGTACTTCC 1131
 DB 7170 GCTGTCACTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTATC 7111
 QY 1132 TATCTGACTAAGACTGCACCAATTTGGTGAACAGCTACTGTTATTTGATGTGCCATAT 1191
 DB 7110 ACTGCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTATC 7051
 QY 1192 CATACTACCAACCACTGTTACCACTGATGAGCAGGAACAATCACTACCAACCAACCACTCGT 1251
 DB 7050 ACTGCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTATC 6991

QY 1252 ACCAATCCAACTGATTCAATTTGACAGAGTGGTGTAAAGTTCCCACTGCAAAATCCAACT 1311
 DB 6990 GTCACTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6931
 QY 1312 GTTAGTACTACTGAATATTTGGTCTCAGTCCCTTTGGTGTACAAACCACTACAGTTACTGCTCT 1371
 DB 6930 ACTATTACTGCTTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6871
 QY 1372 CCAGGTGGTACCGATACTGTTATATCAGAGAGCCACCAACCACTACTGCTCACTACTACT 1431
 DB 6870 GCTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTATT 6811
 QY 1432 GAATATTTGGTCAAAATCCCTTTGCTACTACTACTACTGTTTACTGCTCTCCAGGTGGTACT 1491
 DB 6810 GCTGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6751
 QY 1492 GACTCAGTAATTTATCAGAGAACCAACCAATCAACTGTCTACACACCACTGTAATTTGGTCT 1551
 DB 6750 GTCACTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6691
 QY 1552 CAATCCTTTGCTACTACTACTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1611
 DB 6690 ATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6631
 QY 1612 ATCAGAGAACCTCCAAACCAACTGTCAACCACTGTAATTTGGTCTGCTGCTGCTGCTGCT 1671
 DB 6630 GCTGCTCACTACTATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6571
 QY 1672 ACCAATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1731
 DB 6570 GCTATCACTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6511
 QY 1732 CCAAAACCACTGCTCACTACTACTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1791
 DB 6510 ACTGCTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6451
 QY 1792 GTAATGCTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1851
 DB 6450 GTCACTGCTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6391
 QY 1852 GTCACTACTACTGAGTATTTGGTCTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1911
 DB 6390 GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6331
 QY 1912 CCAAGTGGCACTGATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1971
 DB 6330 ACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6271
 QY 1972 GAATACTGCTGCTCAATCATATGCAACCACTACTACCAATTTACCGCTGCCACTGGTGAACAT 2031
 DB 6270 GTCACTACTATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6211
 QY 2032 GATACCGTTCTTATCAGAGAGCCCAACCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2091
 DB 6210 GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6151
 QY 2092 CAATCATATGCTCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2151
 DB 6150 ACTGCTGCTCACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6091
 QY 2152 ATCAGAGAGCCCAACCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2211
 DB 6090 GTCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6031
 QY 2212 ACAACCACTGTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2271
 DB 6030 GTCGCT 5971
 QY 2272 CCAAAATCCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2331
 DB 5970 GTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5911
 QY 2332 GTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2362

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 03:16:39 ; Search time 3338 seconds
(without alignments)
18369.117 Million cell updates/sec

Title: US-09-715-876-7

Perfect score: 3786

Sequence: 1 atgtctcaacaattacatt.....tgctctgttcatttagta 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_Other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	400.6	10.6	959	17	CNS07DAZ
c 2	353	9.3	893	17	CNS07D66
c 3	272.8	7.2	1050	17	CNS07CMQ
c 4	164.6	4.3	1047	17	CNS07D6L
c 5	144.4	3.8	1011	17	CNS07DAY
c 6	102	2.7	735	17	CNS04NSM
					AL440241 T3 end of
					AL440068 T7 end of
					AL433368 T7 end of
					AL440083 T7 end of
					AL440240 T7 end of
					AL299119 Tetraodon

7	100.4	2.7	773	17	CNS01VTG	AL169549 Tetraodon
8	98.4	2.6	908	17	CNS07DJI	AL40548 T7 end of
9	97	2.6	762	17	CNS01XFI	AL171639 Tetraodon
10	94.8	2.5	824	17	AZ185454	AZ185454 SP_1005_A
11	91.4	2.4	501	17	FR0048173	AL444958 Fugu rubr
12	91.2	2.4	989	17	CNS03HA4	AL197365 Tetraodon
13	89.6	2.4	494	17	FR0048073	AL444858 Fugu rubr
14	82.6	2.2	450	17	FR0025683	AL018519 F.rubripe
15	81.4	2.2	641	13	BM181884	BM181884 fv51b11.y
16	76.2	2.0	694	17	BM355163	BM355163 CH230-81P
c 17	75.6	2.0	530	17	AZ166409	AZ166409 SP_0088_A
18	75.6	2.0	619	17	FR0047601	AL444386 Fugu rubr
19	75	2.0	500	17	B67139	B67139 CpG0015B Cp
c 20	74	2.0	935	17	CNS033D4	AL225985 Tetraodon
21	73.4	1.9	619	17	FR0006944	Z90754 F.rubripes
22	72.6	1.9	644	13	BM605179	BM605179 170006870
c 23	72	1.9	485	17	AZ434703	AZ434703 IM0221C24
24	70.4	1.9	621	17	AZ164800	AZ164800 SP_0076_B
25	69.6	1.8	350	17	BH879665	BH879665 ht47h12.g
26	69.6	1.8	740	10	AV973760	AV973760 AV973760
27	69.2	1.8	788	17	AZ183942	AZ183942 SP_1002_A
28	68.8	1.8	331	17	BH881398	BH881398 hv25c11.b
29	68.4	1.8	933	17	AZ204694	AZ204694 SP_0100_A
c 30	68	1.8	470	17	FR0018463	AL011359 F.rubripe
c 31	67	1.8	810	17	AZ199472	AL199472 SP_1039_B
32	66.4	1.8	358	17	BH777606	BH777606 fzm5013f0
33	65.6	1.7	605	17	AZ640388	AZ640388 IM0502821
c 34	65	1.7	642	17	CI1688	AJ226509 Clona int
c 35	64.2	1.7	778	13	BM170064	BM170064 EST572587
36	64	1.7	914	17	AZ205202	AZ205202 SP_0100_A
c 37	63.4	1.7	695	13	BJ388152	BJ388152 BJ388152
38	62.4	1.6	273	17	FR0047466	AL444251 Fugu rubr
39	62.4	1.6	619	17	FR0013713	AL004959 F.rubripe
40	62	1.6	415	10	BE56771	BE56771 fk95c07.y
c 41	61.6	1.6	979	17	CNS08FDW	AL409242 T7 end of
42	61.4	1.6	550	17	FR0043207	AL130699 Fugu rubr
43	61.4	1.6	718	17	AZ972907	AZ972907 2M0246M24
44	61.2	1.6	619	17	FR0006038	Z89848 F.rubripes
c 45	61.2	1.6	832	13	BM170146	BM170146 EST572669

ALIGNMENTS

RESULT 1	CNS07DAZ	959 bp	DNA	linear	GSS 08-JUL-2001
LOCUS	T3 end of clone BD0AA010H12 of library BD0AA from strain CBS 94 of				
DEFINITION	Candida tropicalis, genomic survey sequence.				
ACCESSION	AL440241				
VERSION	AL440241.1	GI:12223652			
KEYWORDS	GSS.				
SOURCE	Candida tropicalis.				
ORGANISM	Candida tropicalis				
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
AUTHORS	Saccharomycetales; mitosporic Saccharomycetales; Candida.				
	1 (bases 1 to 959)				
	Souci, J.L., Aigle, M., Artiguenave, F., Blandin, G.,				
	Boletín-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,				
	de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,				
	Maupertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Potier, S.,				
	Saurin, W., Tekai, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,				
	Wincker, P., and Weissenbach, J.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of				
	yeast species for molecular evolution studies				
JOURNAL	FEMS Lett. 487 (1), 3-12 (2000)				
MEDLINE	20584711				
PUBMED	11152876				
REFERENCE	2 (bases 1 to 959)				
AUTHORS	Blandin, G., Ozier-Kalogeropoulos, O., Wincker, P., Artiguenave, F. and				
	Dujon, B.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 16. Candida				
	tropicalis				

source

Db
256 ACCCACCAGGTGGCACTGATACCGTTGTTGTCATCGACCCACCAAACTGTCACCA 197

5
 4
 3
 2
 1

```
/clone_11b=...BD
/note="end : T
```

不

yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces* *exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces* *lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia* *angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

```
FEATURES
    source             Location/Qualifiers
        1..1011
            /organism="Candida tropicalis"
            /strain="CBS 94"
            /db_xref="taxon:5482"
            /clone="BD0AA010H12"
            /clone_lib="BD0AA"
            /note="end : 17"
            <652..>988
            /note="similar to P46590 [ Agglutinin-like protein 1
            precursor, ALS1 ] [ Candida albicans]
            1 putative frameshift(s)"
            /evidence=not_experimental

BASE COUNT      336 a 188 c 151 g 293 t    43 others
ORIGIN

Query Match      3.8%; Score 144.4; DB 17; Length 1011;
Best Local Similarity 62.6%; Pred. No. 4.5e-26;
Matches 234; Conservative 4; Mismatches 135; Indels 1; Gaps 1;

QY 7 CAACAAATTTACATGTTTATTCCTATATTTGCAATTTGCAAGTGGCAAGACAATCATCTGTT 66
Db 625 CAACAACTCTCTTATGCTGTGTCATTTGATTTGCTACTTTCAGTTGCTTAAAGAAATTTCCGGT 684

QY 67 GTTTTGTATAGTCTTTTAAATTCATTAATGCTGCTCAATGCTGCTTAATATGCTTTTCAAGGG 126
Db 685 GTGTTTATAGGATTTGAATCTTAACTTGGGATAAGCTGGTAATATGTCATATCAAGT 744

QY 127 CCAGATACCCCAACTTGAATGCTGTTTGGTGGTCTTCTAGATGGTACCAAGTCCCAAT 186
Db 745 CCTCAATATCCAACTGGAATGCCGCTGCTTGAATGGTGGCAACACATCC 804

QY 187 CCAGGGGATACATTCACATTTGAATATGCAATGCTGTTTAAATATACATTTTCAACAACA 246
Db 805 CCAGGTGATACGTTTCACTTTGATCATGCCCTGCTGCTTCAAGTTCACTACGTTGCAVTT 864

QY 247 TCTGTTGATTAATGCTGCGATGCTTAAATATGCTACTTGTCAATTTTAT-TC7GGTGA 305
Db 865 CTGTTGGATCTGACTGCCAATGGKATGACATATGCCACATGTGATCTTCATGCCGGTGA 924

QY 306 AGAATTCACAACCTTTTCTACATTAACATGATGCTGTAACGACGCTTTGAAATCATCCAT 365
Db 925 AGAGTTTACTACTTACTTCTAGTTTACGTTACGTTACTGCTAYAGATGCTTAAAGTCTGTTCA 984

QY 366 TAAGGCATTTGGTA 379
Db 985 GAAGCTATGGGATA 998

RESULT 6
CNS04NSM/c
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence T7 end of clone
123M05 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL299119
VERSION
AL299119.1 GI:8038260
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis
ORGANISM
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
```

```
QY 62 CTGGTGTGTTTTCATAGTTTAAATTCATTAATGCTGCTCAATGCTGCTAAATTA-----TG 115
Db 423 CTGGGATTTTATGTTTTCATTAATTCATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 364

QY 116 CTTTCAAGGCCAGGATACCAACTTGGATGCTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGA 175
Db 363 CATACCAAGGTCCACAAATTCCTACCTGGACTGCAGAACTAACATGCTGTTTGGTGGTGGTGGG 304

QY 176 CCAGTGGCAATCCAGGGGATACATTCACATTAATATGCAATGCTGCTGCTGCTGCTGCTGCTGCT 235
Db 303 AATCTGCTGAACCAAGGTGATACATTCATTAATATGCAATGCTGCTGCTGCTGCTGCTGCTGCTG 244

QY 236 CTTCAACAACTCTGTTGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 295
Db 243 CAATCAACAACTCTGTTGATTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 184

QY 296 ATTCCTGGTGAAGATTCACACTTTTCTACATTAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGA 355
Db 183 ATTCCTGGTGAAGATTTTACTACTTTCTCAAGTTTAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGA 124

QY 356 AATCATCATTAAGCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 415
Db 123 CAACATATACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 64

QY 416 CAGGTTTCAACATGATTTGGAAGATTTCTAAATGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 475
Db 63 CGGGAACCTCTGTTTCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4

QY 476 C 476
Db 3 C 3

RESULT 5
CNS07DAY
LOCUS
DEFINITION
T7 end of clone BD0AA010H12 of library BD0AA from strain CBS 94 of
Candida tropicalis, genomic survey sequence.
ACCESSION
AL440240
VERSION
AL440240.1 GI:12223651
KEYWORDS
GSS.
SOURCE
Candida tropicalis.
ORGANISM
Candida tropicalis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
1 (bases 1 to 1011)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neugeglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
PUBMED
11152876
2 (bases 1 to 1011)
Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
Dujon,B.
Genomic exploration of the hemiascomycetous yeasts: 16. Candida
tropicalis
FEBS Lett. 487 (1), 91-94 (2000)
20584726
PUBMED
11152891
3 (bases 1 to 1011)
Genoscope.
Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web :
This GSS is part of a random genomic sequencing program of thirteen
```

[illegible]

Qy	1486	GGTACTGAC	TCACTAAT	TATCAGAGAAC	CACCAAAATCCAACTGCTCACTACAA	CCGAGTAT	1545
Db	381	AATACTAAT	ACTACTACT	ACTACTTCTACT	TCTACTACTACTACTACT	TCTACTACT	440
Qy	1546	TGGTCTCAAT	CCCTTGGTACTACT	ACTACTACAGTTACT	TGCTCCTCCAGGTGGTACTGACTCA	1605	
Db	441	AATACTACT	ACTACTACT	ACTACTACTACT	TCTACTACTACTACTACTACTACTACT	TCTACTTCT	500
Qy	1606	GTAATTTAT	CAGAGAACTT	CCAAACCCAACTGTCACCA	CACCACTGAAT	1651	
Db	501	ACTACTAA	CCAGGTAA	CCGAGTAAGTAAATTTT	TTTTCATCAACAAATAAAT	546	

RESULT 11	
LOCUS	FR0048173
DEFINITION	501 bp DNA linear GSS 05-JAN-2001
	Fugu rubripes GSS sequence, clone 264522ca9, genomic survey
ACCESSION	FR0048173
VERSION	AL444958.1 GI:12053458
KEYWORDS	GSS: genome survey sequence.
SOURCE	Takifugu rubripes.
ORGANISM	Takifugu rubripes
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
	Tetraodontidae; Takifugu.
REFERENCE	1 (bases 1 to 501)
AUTHORS	Clark, M.S.
TITLE	Direct Submission
JOURNAL	Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource
	Centre Hinxtton, Cambridge, CB10 1SB. UK Email:
	biohelp@hmp.mrc.ac.uk
COMMENT	Vector: pBluescript II KS
	V_type: phagemid
	PRIMER: KS

DESCRIPTORS: One pass dye-terminator sequencing of BAC (pBelosBACII) cloned genomic sequence
The BACs can be obtained from <http://www.incvte.com>.

```

FEATURES             source
Location/Qualifiers
1 501
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone="264E22CA9"
/clone_lib="BAC 264E22"
132 a 161 c 35 g 173 t
BASE COUNT
ORIGIN

```

Query Match	2.4%	Score 91.4;	DB 17;	Length 501;
Best Local Similarity	49.5%;	Pred. No. 1.7e-12;		
Matches 236;	Conservative	0;	Mismatches 241;	Indels 0;
				Gaps 0;

Qy	1123	ACTACTTCCTATCTGACTAAGACTGCACCAATGGTGAAACAGCTACTGTTATTTGTTGAT	1182
Db	12	ACTACTGCTGCTGCTTCTACTGCTACTGCTACTGCTACTGCTACTGCTGCTGCT	71

Qy	1183	GTGCCATATCATCTACTACACACACTGTTACCACTGAATGGACAGGAACAATCTACTACCACC	1242
Db	72	TCTACTACTACTACTACTACTACTGCTACTACTACTACTGCTACTACTACTACTACTACT	131

[illegible][illegible][illegible]

Qy	1423	ACTACTACTGAATATTGGTCACAAATCCTTTGCTACTACTACTACTGTACTGCTCTCCA	1482
Dd	312	ACTACTGCTTCTACTACTACTACTACTACTACTACTACTACTACTGCTGCTTCTACT	371
Qy	1483	GGTGGTACTGACTCGAGTAATTATCAGAGAACCACCAATCCAACTGTCACTACAACGCAG	1542
Dd	372	ACTGCTACTACTACTACTACTACTACTACTGCTACTACTACTAATACTGCTTACT	431
Qy	1543	TATTGGTGCTCAATCCTTTGCTACTACTACTACTACTGCTCTCACGGTGGTACT	1599
Dd	432	ACTACTCTGCTACTACTACTACTACTACTGCTAGTAGTACTACTACTACTACTAGTAGT	488

RESULT 12
CNS02HA4
LOCUS
DEFINITION
 989 bp DNA linear GSS 13-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
139021 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
 AL197365.1 GI:7835515
VERSION
 GSS: genome survey sequence.
KEYWORDS
 Tetraodon nigroviridis.
SOURCE
 Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
REFERENCE
 1 (bases 1 to 989)
 Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
TITLE
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
JOURNAL
 Unpublished
REFERENCE
 2 (bases 1 to 989)

AUTHORS	TITLE
Koest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Flsner, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.	Characterization and repeat analysis of the compact genome of the

JOURNAL REFERENCE	AUTHORS	TITLE	Genome Coordinates	Submission Date	Submitted (12-APR-2000)	Comment
			3 (bases 1 to 989)			

FEATURES	Location/Qualifiers
source	1..989

```

/organism="tetraodon nigrovittatus"
/db_xref="taxon:99883"
/clone="139021"
/clone_lib="G"

```

BASE COUNT	320 a	245 c	122 g	261 t	41 others
ORIGIN					

Best Local Similarity 47.0%; Pred. No. 2 8e-12;
Matches 238; Conservative 9; Mismatches 259; Indels 0; Gaps 0;

[illegible][illegible]

```

QY 3233 CAAGTTTGAACCAAGTATGGGTGAATTCCTGGATTAACTACTTCTACTGAAATGAAG 3292
Db 151 CTAATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 210
QY 3293 CTACAACAACCACTCTACAGAGCTCCATCACCCTGCTGTTCTTCTGCTACTGATGTA 3352
Db 211 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTGCTG 270
QY 3353 CTACTGAACCAACTGATAGAGAACCAACCTACTACTACTACTACTACTACTACTACTACT 3412
Db 271 CTAATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 330
QY 3413 ACAGTGAACCTGGTGTACTACACAAGCTACTTAATGAAATGGTGAATCTCCATCAA 3472
Db 331 CTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 390
QY 3473 CTGATTTAACATCAAGCTTGACACAAGGACCTCAGCATCTCAAGTGTCTAATAGGCAAC 3532
Db 391 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAATA 450
QY 3533 TTGTTTACTAGTGATCTGTTACTGTGGAGCTGTTGCCAGTGTCTTCAAAATGATCAATC 3592
Db 451 ATNNTNNTNGTNGTGGGGTGGTGGTCTCTACTACTACTACTACTACTACTACTAATA 510
QY 3593 ATTCTACTCTGTTTACCAACAGCAAC 3618
Db 511 ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 536

RESULT 13
FR0048073
LOCUS
DEFINITION
Fugu rubripes GSS sequence, clone 263K15BD8, linear GSS 05-JAN-2001
sequence.
AL444858
AL444858.1 GI:12052694
VERSION
GSS: genome survey sequence.
KEYWORDS
Takifugu rubripes
SOURCE
Takifugu rubripes
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
1 (bases 1 to 494)
Clark, M.S.
Direct Submission
Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. UK Email:
biohelp@hgm.mrc.ac.uk
Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of BAC (pBelobACII) cloned
genomic sequence
The BACs can be obtained from http://www.encyte.com.
FEATURES
source
Location/Qualifiers
1..494
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone="263K15BD8"
/clone_lib="BAC 263K15"
BASE COUNT 128 a 163 c 45 g 158 t
ORIGIN
Query Match 2.4%; Score 89.6; DB 17; Length 494;
Best Local Similarity 49.0%; Pred. No. 5e-12;
Matches 239; Conservative 0; Mismatches 249; Indels 0; Gaps 0;

QY 1102 ACAATCTCATATCTTGGTGTGACTACTTCTCTACTACTAGACTCCACCAATTTGGTGAA 1161
Db 6 AGCACTACTACTACTACTCGGGCTTCTACTACTACTACTACTACTACTACTACTACTACTACT 65

```

```

QY 1162 ACAGCTACTGTTATTGTTGATGGCATATCATACTACCACAACTGTTTACCAGTGAATGG 1221
Db 66 AGCACTACTGCTGGCGTCGACTACTGGACGAGCACTGCTACTACTACTGCGGCTGCTTCT 125
QY 1222 ACAGGAAACAATCACTACCACCACCAACTCGTACCAATCCAACCTGATTCAATTGACACAGTG 1281
Db 126 ACTACTGCTACTACTACTAGGAGCACTGCTACTACTACTACTACTACTACTACTACTACT 185
QY 1282 GTGGTACAAGTTCCTACTGCGCAATCCAACCTGTTAGTACTCAATATTGGTCTCAGTCC 1341
Db 186 ACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 245
QY 1342 TTTGCTACAACCACTACAGTTTACTGCTCCTCCAGGTGGTACCGATACTGTGATATCAGA 1401
Db 246 ACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 305
QY 1402 GAGCCACCAACCACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACT 1461
Db 306 GCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 365
QY 1462 ACTACTGTTACTGCTCCTCCAGGTGCTACTGCTACTAGTAATATATCAGAACCCACCAAT 1521
Db 366 ACTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 425
QY 1522 CCAACTGTCTACACACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACTACTACT 1581
Db 426 ACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 485
QY 1582 GCTCCTCC 1589
Db 486 GCTACTAC 493

RESULT 14
FR0025683
LOCUS
DEFINITION
F. rubripes GSS sequence, clone 154E17ac12, linear GSS 10-DEC-1997
sequence.
AL018519
AL018519.1 GI:2684887
VERSION
GSS: genome survey sequence.
KEYWORDS
Takifugu rubripes
SOURCE
Takifugu rubripes
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Takifugu.
1 (bases 1 to 450)
Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrania, Y.,
Williams, G. and Brenner, S.
Direct Submission
Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgm.mrc.ac.uk
Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR:
One pass dye-terminator sequencing of cosmid cloned genomic
sequence.
FEATURES
source
Location/Qualifiers
1..450
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone="154E17ac12"
/clone_lib="cosmid 154E17"
BASE COUNT 96 a 140 c 64 g 144 t
ORIGIN
Query Match 2.2%; Score 82.6; DB 17; Length 450;
Best Local Similarity 49.3%; Pred. No. 3.1e-10;
Matches 208; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 1276 ACAGTGGTGTACAGTTCCTCCAGTCCCAATCCAACTGTTAGTACTACTGAATATGCTCT 1335

```

Db 29 ACTGGAGACNTCAATACCTCAGTGGCCTCTAACTCTACTAATGCTACTACTACT 88
 QY 1336 CAGTCCTTTGTCAACACACTACAGTACTCTCTCCAGTGGGTACCGACTGTGATT 1395
 Db 89 ACTGCTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 148
 QY 1396 ATCAGAGGCCACCAACACTACTACTACTACTACTACTACTACTACTACTACT 1455
 Db 149 GCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 208
 QY 1456 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1515
 Db 209 ACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 268
 QY 1516 CCAATCAACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACT 1575
 Db 269 ACTGCTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACT 328
 QY 1576 GTTACTGCTCTCAGGTGGTACTACTACTACTACTACTACTACTACTACTACTACT 1635
 Db 329 GCTACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACT 388
 QY 1636 GTCACCACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACT 1695
 Db 389 GCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 448
 QY 1696 CC 1697
 Db 449 NC 450

RESULT 15
 BM181884
 LOCUS
 DEFINITION f51bl.1 Sugano SJD adult male Danio rerio cDNA clone 5412044 5' similar to contains element TARI repetitive element ;, mRNA

ACCESSION
 VERSION BM181884.1 GI:17512842
 KEYWORDS
 SOURCE zebrafish.

ORGANISM
 Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE
 AUTHORS Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

TITLE Washu zebrafish EST Project 1998
 JOURNAL Unpublished (1998)
 COMMENT Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: zbrafish@wustl.edu
 Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
 Seq primer: T3 Er from Amersham
 High quality sequence stop: 527.

FEATURES
 source
 1..641
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone="5412044"
 /clone_lib="Sugano SJD adult male"
 /sex="male"

/tissue_type="whole body"
 /dev_stage="adult"
 /lab_host="DH10B (phage resistant)"
 /note="vector: pME18S-FL3; Site 1: DraIII (CACCATGNG); Site 2: DraIII (CACTGTGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGGGCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGNG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed and donated by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTGCTCTAAAGCTGCG and 3' end primer CGACCTCGAGCTCGAGCACA."

BASE COUNT 218 a 234 c 37 g 152 t
 ORIGIN
 Query Match 2.2%; Score 81.4; DB 13;
 Best Local Similarity 48.1%; Pred. No. 7.7e-10;
 Matches 297; Conservative 0; Mismatches 311; Indels 9; Gaps 2;
 QY 2661 ATCTCAATGTAAGTACTACTGTTCTTCTGATAATGTTTCAAAATCAGAGTATCACTTTC 2720
 Db 34 ATCTCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 93
 QY 2721 AACTGATGTTAGTGTCTACTATCTTCTGATAATGTTTCAAAATCAGAGTATCACTTTC 2780
 Db 94 TACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 153
 QY 2781 AACGAAATCTTCTGTACAACTATTCAAACTACTTCAAAACCACTTATCATCTTTCAGTGAC 2840
 Db 154 AACACACAACTTCTCCAGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 213
 QY 2841 ATCAATGACTCAGTTGTTCTCAATTCCTCAAGTGTTCAGAAAAGTGAAGTAAAGTTACATT 2900
 Db 214 AACAACTTCTCCAGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 270
 QY 2901 TACAGCAATGGAGACAAACCAAGTGTCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 2960
 Db 271 AACTTCAACAAACAACTTCTCCAGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 330
 QY 2961 TGAATTTGAACAAACCACTTCTTAAAGTGTTCAGCACTGCTGCTGCTTCTTCTTAACTACTGA 3020
 Db 331 AGCTACTCACTTCTCCAAACAAACAACTTCTCCAGCTACTGCTACTGCTACTGCTACTGCTACTGCT 390
 QY 3021 TTTGACTAGTGAACCAACAAATATACAGAGAAACAACTACTGCTACTGCTACTGCTACTGCTACTGCT 3080
 Db 391 AACAACTTCTCCAGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 450
 QY 3081 CTCCATCACTGAAGTATACCACTATCCACATCTCAACCTACAGGTGATATGAGAGACAATCTTTC 3140
 Db 451 TTCAACAAC-----AACAACTTCTCCAGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 504
 QY 3141 ATCAACCAATCCAGTTTCCAACTGTGGCAACAGTACTTCTAGCTCTGCAAGTGAAGAAGA 3200
 Db 505 TCCAGCTACAATCTTCTCCAGCAACCAACATCACTTCTCCAGCTACTGCTACTGCTACTGCTACTGCTACTGCT 564
 QY 3201 CAACAAAAGCGTGTCTCATGAATCAGCTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3260
 Db 565 AACAACTTCTCCAGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 624
 QY 3261 TTCTGGATTACTACTT 3277
 Db 625 TTCTCCAGCAACACAT 641

Search completed: June 20, 2003, 06:13:39
 Job time : 3348 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 03:17:58 ; Search time 128 Seconds
(without alignments)
9070.924 Million cell updates/sec

Title: US-09-715-876-7
Perfect score: 3786
Sequence: 1 atgttcaacaattacatt.....tgtctgttcatttagtga 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

*Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1069.4	28.2	1071	1 US-08-357-962-1	Sequence 1, Appli
2	1069.4	28.2	1071	1 US-08-878-106-1	Sequence 1, Appli
3	348.2	9.2	425	1 US-08-357-962-5	Sequence 5, Appli
4	348.2	9.2	425	1 US-08-878-106-5	Sequence 5, Appli
5	87	2.3	5511	3 US-08-928-361B-2	Sequence 2, Appli
6	87	2.3	7334	3 US-08-928-361B-1	Sequence 1, Appli
7	86.2	2.3	5163	3 US-08-700-651-1	Sequence 1, Appli
8	86.2	2.3	5163	3 US-08-928-361B-4	Sequence 4, Appli
9	86.2	2.3	5318	3 US-08-700-651-2	Sequence 2, Appli
10	86.2	2.3	5318	3 US-08-928-361B-3	Sequence 3, Appli
11	85.2	2.3	108	1 US-08-357-962-2	Sequence 2, Appli
12	85.2	2.3	108	1 US-08-878-106-2	Sequence 2, Appli
13	82.4	2.2	100	1 US-08-145-705A-32	Sequence 32, Appli
14	82.4	2.2	100	1 US-08-145-705A-34	Sequence 34, Appli
15	76	2.0	100	1 US-08-145-705A-33	Sequence 33, Appli
16	68	1.8	100	1 US-08-145-705A-36	Sequence 36, Appli
17	58	1.5	3168	4 US-09-165-239A-3	Sequence 3, Appli
18	57.2	1.5	100	1 US-08-145-705A-37	Sequence 37, Appli
19	57	1.5	100	1 US-08-145-705A-35	Sequence 35, Appli
20	54.2	1.4	3489	2 US-08-728-323A-1	Sequence 1, Appli
21	54.2	1.4	3489	4 US-09-298-568-1	Sequence 1, Appli
22	54.2	1.4	32207	2 US-08-770-379-20	Sequence 20, Appli
23	54.2	1.4	32207	4 US-08-757-669A-20	Sequence 20, Appli
24	54.2	1.4	32207	4 US-09-230-371A-20	Sequence 20, Appli
25	50.2	1.3	4197	2 US-08-682-517-7	Sequence 7, Appli
26	50.2	1.3	4197	2 US-08-682-517-8	Sequence 8, Appli
27	49	1.3	3337	1 US-08-072-610-1	Sequence 1, Appli

c 28	49	1.3	3337	2 US-08-719-822B-1	Sequence 1, Appli
c 29	49	1.3	3337	4 US-09-092-458-1	Sequence 1, Appli
c 30	48.8	1.3	3183	4 US-08-911-393-1	Sequence 1, Appli
c 31	47.8	1.3	2824	4 US-07-757-022B-13	Sequence 13, Appli
c 32	47.8	1.3	3066	4 US-07-757-022B-83	Sequence 83, Appli
c 33	47.8	1.3	3117	4 US-07-757-022B-73	Sequence 73, Appli
c 34	47.8	1.3	3148	4 US-07-757-022B-57	Sequence 57, Appli
c 35	47.8	1.3	3420	4 US-07-757-022B-103	Sequence 103, Appli
c 36	47.8	1.3	3813	4 US-07-757-022B-43	Sequence 43, Appli
c 37	47.8	1.3	3936	4 US-07-757-022B-41	Sequence 41, Appli
c 38	47.8	1.3	3942	4 US-07-757-022B-141	Sequence 141, Appli
c 39	47.8	1.3	3945	4 US-07-757-022B-49	Sequence 49, Appli
c 40	47.8	1.3	3963	4 US-07-757-022B-45	Sequence 45, Appli
c 41	47.8	1.3	3963	4 US-07-757-022B-59	Sequence 59, Appli
c 42	47.8	1.3	4065	4 US-07-757-022B-47	Sequence 47, Appli
c 43	47.8	1.3	4086	4 US-07-757-022B-39	Sequence 39, Appli
c 44	47.8	1.3	4092	4 US-07-757-022B-51	Sequence 51, Appli
c 45	47.8	1.3	4215	4 US-07-757-022B-61	Sequence 61, Appli

ALIGNMENTS

RESULT 1
US-08-357-962-1
; Sequence 1, Application US/08357962
; Patent No. 5668263
; GENERAL INFORMATION:
; APPLICANT: Hoyer, Lois
; APPLICANT: Livi, George
; APPLICANT: Shatzman, Allan
; TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION NUMBER:
; APPLICATION NUMBER: US/08/357,962
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert H
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P50278
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5019
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1071 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-357-962-1

Query Match 28.2%; Score 1069.4; DB 1; Length 1071;
 Best Local Similarity 99.9%; Pred. No. 5.8e-252;
 Matches 1070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1300 CCAATCCCACTGTTAGTACTACTGAATATTGGTCTCAGTCTCTTGGCTACACCACTACTA 1359
 DB 1 CCAATCCCACTGTTAGTACTACTGAATATTGGTCTCAGTCTCTTGGCTACACCACTACTA 60

QY 1360 GTTACTGCTCTCCAGGTGGTACCGATAGTATTATCAGAGAGCCACCAACCATACT 1419
 DB 61 GTTACTGCTCTCCAGGTGGTACCGATAGTATTATCAGAGAGCCACCAACCATACT 120

QY 1420 GTCACTACTACTGAATATTGGTACCAATCTTTGGTCTACTACTACTACTACTACTCTCT 1479
 DB 121 GTCACTACTACTGAATATTGGTACCAATCTTTGGTCTACTACTACTACTACTACTCTCT 180

QY 1480 CCAGGTGGTACTGACTCAGTAATATTATCAGAGAACCAACCAATCCAACTGTCTACTACAAC 1539
 DB 181 CCAGGTGGTACTGACTCAGTAATATTATCAGAGAACCAACCAATCCAACTGTCTACTACAAC 240

QY 1540 GAGTATTGGTCTCAATCTCTTGGTCTACTACTACTAGTATTACTGCTCTCCAGGTGGTACT 1599
 DB 241 GAGTATTGGTCTCAATCTCTTGGTCTACTACTACTAGTATTACTGCTCTCCAGGTGGTACT 300

QY 1600 GACTCAGTAATATTATCAGAGAACCTCCAAACCACTGTCCACCACTGAATATTGGTCC 1659
 DB 301 GACTCAGTAATATTATCAGAGAACCTCCAAACCACTGTCCACCACTGAATATTGGTCC 360

QY 1660 CAATCTTACGACCACTACTGACTGCTGCTCTCCAGGAGGCACTGACTCAGTAAT 1719
 DB 361 CAATCTTACGACCACTACTGACTGCTGCTCTCCAGGAGGCACTGACTCAGTAAT 420

QY 1720 ATCAGAGAACCAACCACTGCTACTACTACTACTGTAATCTGGTCACAATCATATGCC 1779
 DB 421 ATCAGAGAACCAACCACTGCTACTACTACTACTGTAATCTGGTCACAATCATATGCC 480

QY 1780 ACCTACTACTGTTACTGCTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1839
 DB 481 ACCTACTACTGTTACTGCTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

QY 1840 CCAATCCCACTGTTAGTACTACTGAATATTGGTCTCAGTCTCTTGGCTACACCACT 1899
 DB 541 CCAATCCCACTGTTAGTACTACTGAATATTGGTCTCAGTCTCTTGGCTACACCACT 600

QY 1900 GTAATCTGGTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1959
 DB 601 GTAATCTGGTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 560

QY 1960 GTCACTACTACTGAATATTGGTCTCAATCATATGCAACCACTACTACTACTACTACTACT 2019
 DB 661 GTCACTACTACTGAATATTGGTCTCAATCATATGCAACCACTACTACTACTACTACTACTACT 720

QY 2020 CTTGGTGAATGATACCGTCTTCTTATCAGAGAGCCACCAACCATACTGCTACTACTACT 2079
 DB 721 CTTGGTGAATGATACCGTCTTCTTATCAGAGAGCCACCAACCATACTGCTACTACTACTACT 780

QY 2080 GAATACTGCTCTCAATCATATGCTCAACCACTGTTACTGCAACCACTGCTGCTGCTGCTGCT 2139
 DB 781 GAATACTGCTCTCAATCATATGCTCAACCACTGTTACTGCAACCACTGCTGCTGCTGCTGCT 840

QY 2140 GATACGGTCTTATCAGAGAGCCACCAACCATACTGCTACTACTACTACTACTACTACTACT 2199
 DB 841 GATACGGTCTTATCAGAGAGCCACCAACCATACTGCTACTACTACTACTACTACTACTACT 900

QY 2200 CAATCATATGCTACACCACTGTTACTGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2259
 DB 901 CAATCATATGCTACACCACTGTTACTGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960

QY 2260 ATTAGAGAGCCACCAATCCAACTGTTACTACTACTACTACTACTACTACTACTACTACTACT 2319
 DB 961 ATTAGAGAGCCACCAATCCAACTGTTACTACTACTACTACTACTACTACTACTACTACTACT 1020

QY 2320 ACAACCACTGTTACTGCTCTCTCCAGGTGGTACTGACTGCTGCTGCTGCTGCTGCTGCTGCT 2370

DB 1021 ACAACCACTGTTAGTACTACTGCTCTCCAGGTGGTACTGACTGCTGATTATC 1071

RESULT 2
 US-08-878-106-1
 ; Sequence 1, Application US/08878106
 ; Patent No. 5817466
 ; GENERAL INFORMATION:
 ; APPLICANT: Hoyer, Lois
 ; APPLICANT: Livi, George
 ; APPLICANT: Shatzman, Allan
 ; TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SmithKline Beecham Corporation
 ; STREET: 709 Swedeland Road
 ; CITY: King of Prussia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19406-0939
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSEQ Version 1.5
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/878,106
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/357,962
 ; FILING DATE: 16-DEC-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jervis, Herbert H
 ; REGISTRATION NUMBER: 31,171
 ; REFERENCE/DOCKET NUMBER: P50278
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 610-270-5019
 ; TELEFAX: 610-270-5090
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1071 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE:
 ; ORIGINAL SOURCE:
 ; US-08-878-106-1

Query Match 28.2%; Score 1069.4; DB 1; Length 1071;
 Best Local Similarity 99.9%; Pred. No. 5.8e-252;
 Matches 1070; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1300 CCAATCCCACTGTTAGTACTACTGAATATTGGTCTCAGTCTCTTGGCTACACCACTACTA 1359
 DB 1 CCAATCCCACTGTTAGTACTACTGAATATTGGTCTCAGTCTCTTGGCTACACCACTACTA 60

QY 1360 GTTACTGCTCTCCAGGTGGTACCGATAGTATTATCAGAGAGCCACCAACCATACT 1419
 DB 61 GTTACTGCTCTCCAGGTGGTACCGATAGTATTATCAGAGAGCCACCAACCATACT 120

QY 1420 GTCACTACTACTGAATATTGGTACCAATCTTTGGTCTACTACTACTACTACTACTCTCT 1479
 DB 121 GTCACTACTACTGAATATTGGTACCAATCTTTGGTCTACTACTACTACTACTACTCTCT 180

QY 1480 CCAGGTGGTACTGACTCAGTAATATTATCAGAGAACCAACCAATCCAACTGTCTACTACAAC 1539
 DB 181 CCAGGTGGTACTGACTCAGTAATATTATCAGAGAACCAACCAATCCAACTGTCTACTACAAC 240

Query Match 2.3%; Score 86.2; DB 3; Length 5163;
Best Local Similarity 47.3%; Pred. No. 1.4e-11;
Matches 311; Conservative 0; Mismatches 328; Indels 18; Gaps 1;
QY 1406 CACCAACATCTGCTCAGTCTACTGATATTTGGTCACAAATCCTTTGCTACTACTACTA 1465
DB 588 CAACAATCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 647
QY 1466 CTGTTACTGCTCTCCAGTGGTACTGCTAGTATTTATCAGAGAACCAACAAATCCAA 1525
DB 648 CTACACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 707
QY 1526 CTGCTACTACAACCGAGTATTTGGTCTCAATCTTTGCTACTACTACTACTACTACTGTC 1585
DB 708 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 767
QY 1586 CTCCAGGTGGTACTGCTAGTATTTATCAGAGAACCTCCAAACCAACTGTACACACA 1645
DB 768 CTACTACTACTACAACCAACAACAACAACAACAACAACAACAACAACAACAACAACA 827
QY 1646 CTGAATATTGGTCCCAATCTTACGCAACCAACAACAACAACAACAACAACAACAACA 1705
DB 828 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 887
QY 1706 CTGACTCAGTATTTATCAGAGAACCAACAACAACAACAACAACAACAACAACAACA 1765
DB 888 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 947
QY 1766 CACAATCATATGCGACCACTAGCTAGTATTTATCAGAGAACCTCCAAACCAACTGTAC 1825
DB 948 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1007
QY 1826 TCATTAGAGAGCCCAACCAACAACAACAACAACAACAACAACAACAACAACAACAACA 1867
DB 1008 CAAGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1067
QY 1868 ATTGGTCTCAATGTTTGTACTACCAACAACAACAACAACAACAACAACAACAACAACA 1927
DB 1068 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1127
QY 1928 CTGTTATCATTTAGGACCAACAACAACAACAACAACAACAACAACAACAACAACAACA 1987
DB 1128 CTACCAAGAAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1187
QY 1988 CATATGCAACCACTACTACCATTTACCGCTCCACCTGGTGAACCTGATACCGTTCTTA 2044
DB 1188 CAACAATCACTGCTGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1244

RESULT 8
US-08-928-361B-4
Sequence 4, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIKSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verny, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 5163 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-928-361B-4
Query Match 2.3%; Score 86.2; DB 3; Length 5163;
Best Local Similarity 47.3%; Pred. No. 1.4e-11;
Matches 311; Conservative 0; Mismatches 328; Indels 18; Gaps 1;
QY 1406 CACCAACATCTGCTCAGTCTACTGATATTTGGTCACAAATCCTTTGCTACTACTACTA 1465
DB 587 CAACAATCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 646
QY 1466 CTGTTACTGCTCTCCAGTGGTACTGCTAGTATTTATCAGAGAACCAACAAATCCAA 1525
DB 647 CTACACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 706
QY 1526 CTGCTACTACAACCGAGTATTTGGTCTCAATCTTTGCTACTACTACTACTACTACTGTC 1585
DB 707 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 766
QY 1586 CTCCAGGTGGTACTGCTAGTATTTATCAGAGAACCTCCAAACCAACTGTACACCA 1645
DB 767 CTACTACTACTACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 826
QY 1646 CTGAATATTGGTCCCAATCTTACGCAACCAACAACAACAACAACAACAACAACAACA 1705
DB 827 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 886
QY 1706 CTGACTCAGTATTTATCAGAGAACCAACAACAACAACAACAACAACAACAACAACA 1765
DB 887 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 946
QY 1766 CACAATCATATGCGACCACTAGCTAGTATTTATCAGAGAACCAACCACTGGTGTAC 1825
DB 947 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1006
QY 1826 TCATTAGAGAGCCCAACCAACAACAACAACAACAACAACAACAACAACAACAACAACA 1867
DB 1007 CAAGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1066
QY 1868 ATTGGTCTCAATGTTTGTACTACCAACAACAACAACAACAACAACAACAACAACAACA 1927
DB 1067 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1126
QY 1928 CTGTTATCATTTAGGACCAACAACAACAACAACAACAACAACAACAACAACAACAACA 1987
DB 1127 CTACCAAGAAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1186
QY 1988 CATATGCAACCACTACTACCATTTACCGCTCCACCTGGTGAACCTGATACCGTTCTTA 2044
DB 1187 CAACAATCACTGCTGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1243
RESULT 9

```

US-08-700-651-2
; Sequence 2, Application US/08700651B
; Patent No. 6015862
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JTRI
; TITLE OF INVENTION: VACCINES, ANTIBIOTICS, AND ADJUVANTS FOR PROPHYLAXIS
; TITLE OF INVENTION: FOR PROPHYLAXIS
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700-651-2
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415, 08/416, 08/417, 08/418, 08/419, 08/420, 08/421, 08/422, 08/423, 08/424, 08/425, 08/426, 08/427, 08/428, 08/429, 08/430, 08/431, 08/432, 08/433, 08/434, 08/435, 08/436, 08/437, 08/438, 08/439, 08/440, 08/441, 08/442, 08/443, 08/444, 08/445, 08/446, 08/447, 08/448, 08/449, 08/450, 08/451, 08/452, 08/453, 08/454, 08/455, 08/456, 08/457, 08/458, 08/459, 08/460, 08/461, 08/462, 08/463, 08/464, 08/465, 08/466, 08/467, 08/468, 08/469, 08/470, 08/471, 08/472, 08/473, 08/474, 08/475, 08/476, 08/477, 08/478, 08/479, 08/480, 08/481, 08/482, 08/483, 08/484, 08/485, 08/486, 08/487, 08/488, 08/489, 08/490, 08/491, 08/492, 08/493, 08/494, 08/495, 08/496, 08/497, 08/498, 08/499, 08/500, 08/501, 08/502, 08/503, 08/504, 08/505, 08/506, 08/507, 08/508, 08/509, 08/510, 08/511, 08/512, 08/513, 08/514, 08/515, 08/516, 08/517, 08/518, 08/519, 08/520, 08/521, 08/522, 08/523, 08/524, 08/525, 08/526, 08/527, 08/528, 08/529, 08/530, 08/531, 08/532, 08/533, 08/534, 08/535, 08/536, 08/537, 08/538, 08/539, 08/540, 08/541, 08/542, 08/543, 08/544, 08/545, 08/546, 08/547, 08/548, 08/549, 08/550, 08/551, 08/552, 08/553, 08/554, 08/555, 08/556, 08/557, 08/558, 08/559, 08/560, 08/561, 08/562, 08/563, 08/564, 08/565, 08/566, 08/567, 08/568, 08/569, 08/570, 08/571, 08/572, 08/573, 08/574, 08/575, 08/576, 08/577, 08/578, 08/579, 08/580, 08/581, 08/582, 08/583, 08/584, 08/585, 08/586, 08/587, 08/588, 08/589, 08/590, 08/591, 08/592, 08/593, 08/594, 08/595, 08/596, 08/597, 08/598, 08/599, 08/600, 08/601, 08/602, 08/603, 08/604, 08/605, 08/606, 08/607, 08/608, 08/609, 08/610, 08/611, 08/612, 08/613, 08/614, 08/615, 08/616, 08/617, 08/618, 08/619, 08/620, 08/621, 08/622, 08/623, 08/624, 08/625, 08/626, 08/627, 08/628, 08/629, 08/630, 08/631, 08/632, 08/633, 08/634, 08/635, 08/636, 08/637, 08/638, 08/639, 08/640, 08/641, 08/642, 08/643, 08/644, 08/645, 08/646, 08/647, 08/648, 08/649, 08/650, 08/651, 08/652, 08/653, 08/654, 08/655, 08/656, 08/657, 08/658, 08/659, 08/660, 08/661, 08/662, 08/663, 08/664, 08/665, 08/666, 08/667, 08/668, 08/669, 08/670, 08/671, 08/672, 08/673, 08/674, 08/675, 08/676, 08/677, 08/678, 08/679, 08/680, 08/681, 08/682, 08/683, 08/684, 08/685, 08/686, 08/687, 08/688, 08/689, 08/690, 08/691, 08/692, 08/693, 08/694, 08/695, 08/696, 08/697, 08/698, 08/699, 08/700
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5318
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-08-700-651-2

```

Query Match	2.3%	Score 86.2;	DB 3;	Length 5318;
Best Local Similarity	47.3%;	Pred. No. 1.4e-11;		
Matches 311; Conservative	0;	Mismatches 328;	Indels 18;	Gaps 1;
Qy	1406	CACCAAAACCATACTGTCACACTACTACTACTAGTAATTTGGTGCACAATCCTTTGGCTACTACTACTA	1465	
Db	588	CAACAACACTACTACTACTACTACTCGCACACAACAACAACAACGACACAACAACAACAACAACA	647	
Qy	1466	CTGTGTACTGCTCTCCAGGTGGTGACTGACATCATGTAATTATTCAGAGAACCACCAAATGCCAA	1525	
Db	648	CTACAACACTACTACCACTACTACTACGACAACAACAACAACAACAACAACAACAACAACAACA	707	
Qy	1526	CTGTCACTAACACCGAGTATTTGGTCTCAATCTCTTTGCTACTACTACTACTACTACTACTACTGCTC	1585	
Db	708	CAACAACAACAACAACAACAACAACAACAACGACTACTACTACTACTACTACTACTACTACTACTA	767	
Qy	1586	CTCCAGGTGGTAGTACTGACTCAGTAAATTATTCAGAGAAGCTCCAACCCAACTGTGCACCACA	1645	
Db	768	CTACTACTACTACAACGACAACAACAACACTACAACCACAACACTACAACGACAACAACACTACA	827	
Qy	1646	CTGAATATTGGTCCCCTTCTTACGCAACCAACAACACTACTGTGACTGCTCTCCAGGAGGCA	1705	
Db	828	CAACAACAACAACAACAACAACACTACACACCAACACTACAACCAACAACACTACAACCAACA	887	
Qy	1706	CTGACTCAGTAAATTATTCAGAGAACCACAACCAACACTGTCACTACTACTTGAATATCTGGT	1765	
Db	888	CAACCACAACCAACCAACCAACTACCAAGAAAACCAACAACAACAACACTACAACAACAACA	947	
Qy	1766	CACAATCATATGCCACCACTACCACACTGTACTGCAACCAACAGGTGGTACTGCACACTGTTA	1825	
Db	948	CAACAACAACAACACTACTACTACAACCAACCAACAACAACAACAACAACAACAACAACAACA	1007	
Qy	1826	TCAATTAGAGAGCCACCAAAACCAACT-----GTCACTACTACTGAGT	1867	
Db	1008	CCAAGAAACCAACAACACTACTACTACTACTACCACAACAACAACAACACTACTACTACTACCA	1067	
Qy	1868	ATTGGTCTCAATCGTTTGTACTACCAACAACACTGTACTGCTGCAACCAAGTGGCACTGATA	1927	
Db	1068	CAACAACAACAACACTACTACTACTACAACAACAACAACAACAACAACAACAACAACAACAACA	1127	
Qy	1928	CTGTTTATCAITGGGGAACCAACCAACCACTGTCAACCACTCTGTAATCTGCTGCTCAAT	1987	
Db	1128	CTACCGAACAACAACAACAACAACAACAACACTACTACTACTACTACAACAAGAACAACA	1187	
Qy	1988	CATATGCACCACTACTACATTTACCGCTCCACCTGGTGAACCTGATACCGTTCTTTA	2044	
Db	1188	CAACAACACTCACTGGCACAAACAACACTACTACTCTGAACTGGAGAGTGAATTA	1244	

```

RESULT 10
US-08-928-361B-3
: Sequence 3, Application US/08928361B
: Patent No. 6071518
: GENERAL INFORMATION:
: APPLICANT: Petersen, Carolyn
: TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS.
: TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
: TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
: TITLE OF INVENTION: SPECIES INFECTIONS
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PETERS, VERNY, JONES & BIKSA
: STREET: 385 Sherman Avenue, Suite 6
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306-1840
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/928,361B
: FILING DATE: 12-SEP-1997
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/026,062
: FILING DATE: 13-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Verny, Hana
: REGISTRATION NUMBER: 30,518
: REFERENCE/DOCKET NUMBER: 480.76-1(HV)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-324-1677
: TELEFAX: 650-324-1678
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5318 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-928-361B-3

```

Query Match	2.3%	Score 86.2	DB 3	Length 5318
Best Local Similarity	47.3%	Pred. No. 1.4e-11		
Matches 311	Conservative 0	Mismatches 328	Indels 18	Gaps 1
Qy	1406	CACCAACCATAGTGCACCTACTACTAGTAATATTGGTCACAAATCCTTTGCTACTACTACTACTA	1465	
Db	587	CAACAACCTACTACTACTACTACTACGACACACAAACGACAACGACAACACACAACTA	646	
Qy	1466	CTGTCTACTGCTCTCCAGGTGGTACTGACTCAGTAATTTACAGAGAACCAACCAATCCAA	1525	
Db	647	CTACAACCTACTACCACCTACTACTACGACACACACACACACACACACACACACAA	706	
Qy	1526	CTGTCACTACAACGGAGTATTGGTCTCAATCTTTGCTACTACTACTACTACTACTGCTC	1585	
Db	707	CAACAACAACAACAACAACAACAACACGACTACTACTACTACTACTACTACTACTACTA	766	
Qy	1586	CTCCAGGTGGTACTGACTCAGTAATTTACAGAGAACTCCAAACCCAACTGTCCACCACCA	1645	
Db	767	CTACTACTACTACAACCACAACAACCTACAACCAACACTACAACCACAACAACCTACAACCA	826	
Qy	1646	CTGAATATTGGTCCCAATCTTTAGCGAACCAACAACTACTGTGCTCTCTCCAGGAGGCA	1705	
Db	827	CAACACAACAACCAACAACAACCTACACCAACACTACAACCAACAACCTACAACCAACAA	886	
Qy	1706	CTGACTCAGTAATTTATCAGAGAACCAACCAACCAACACTGTCTACTACTACTGAACTACCT	1765	
Db	887	CAACCAACAACCAACCAACCAACCTACCAAGAAACCAACAACAACAACCTACAACCAACAA	946	

	Best Local Similarity	60.2%;	Pred. No. 5.4e-12;	Mismatches	2;	Indels	0;	Gaps	0;
	Matches	65;	Conservative	41;					
Qy	1300	CCAAATCCAACTGTTAGTACTACTGTAATATGGTCTCAGTCCTTTGCTACAAACCACTACA	1359						
Db	1	CCAAAYCMHACWGTYSYACHACYGARTATGGTCNCARTCNTWYGGCHACHACHACHACH	60						
Qy	1360	GTTACTGCTCCTCCAGGTGTCACCGTACTGTGATTATCATGAGAGCCA	1407						
Db	61	RTDACYGSCWCCWGRGHACYGAYWCHGTDMTATYAGRCAGCW	108						

RESULT 12

US-08-878-106-2
: Sequence 2, Application US/08878106
: Patent No. 5817466
: GENERAL INFORMATION:
: APPLICANT: Hoyer, Lois
: APPLICANT: Livl, George
: APPLICANT: Shatzman, Allan
: TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Smithkline Beecham Corporation
: STREET: 709 Swedeland Road
: CITY: King of Prussia
: STATE: PA
: COUNTRY: USA
: ZIP: 19406-0939
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/878,106
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/357,962
: FILING DATE: 16-DEC-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Jarvis, Herbert H.
: REGISTRATION NUMBER: 31,171
: REFERENCE/DOCKET NUMBER: P50278
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-270-5019
: TELEFAX: 610-270-5090
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 108 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: NO
: ORIGINAL SOURCE:
US-08-878-106-2

	Query Match	2.3%	Score 85.2	DB 1	Length 108
	Best Local Similarity	60.2%	Pred. No. 5.4e-12		
	Matches 65	Conservative 41	Mismatches 2	Indels 0	Gaps 0
Qy	1300	CCAAATCCAACTCTTACTACTACTGAATATTTGCTCAGTCCTTCTCTCAAAACCACTACA	1359		
Db	1	CCAAATCMHACWGTAYAYACHACYGARTATVTGGTCNCARTCTNTWYGHACHACHACHACH	60		
Qy	1360	GTTACTGCTCCTCCAGTGTGTCACGACTACTGTGATTATCAGAGAGCCA	1407		
Db	61	RTDACYGSMCCWCCWRGWRGHACYGAWCHGTDMTATYAGRCARCW	108		

RESULT 13
US-08-145-705A-32/c
; Sequence 32, Application US/08145705A
; Patent No. 5489513

```

: GENERAL INFORMATION:
: APPLICANT: Springer, Wolfgang; Plempel, Manfred;
: APPLICANT: L. Bering, Antonius
: TITLE OF INVENTION: SPECIFIC GENE PROBES AND
: TITLE OF INVENTION: PROCESSES FOR THE DIAGNOSTIC
: TITLE OF INVENTION: INVESTIGATION OF CANDIDA
: TITLE OF INVENTION: ALBAINS
: NUMBER OF SEQUENCES: 44
:

```

```

:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: SPRUNG HORN KRAMER & WOODS
:
: STREET: 660 White Plains Road
:
: CITY: Tarrytown
:
: STATE: New York
:
: COUNTRY: U.S.A.
:
: ZIP: 10591-5144
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Diskette, 3.5 inch, 1.4 MB storage
:
: COMPUTER: NEC Powermate I Plus
:
: OPERATING SYSTEM: DOS
:
: SOFTWARE: Wordperfect 5.1
:

```

Query Match	2.2%	Score 82.4;	DB 1;	Length 100;
Best Local Similarity	89.0%	Pred. No. 2.5e-11;		
Matches 89;	Conservative	0;	Mismatches 11;	Indels 0;
				Gaps 0;

QY 1230 AATCACTACCACCACAACCTCGTACCAATCCAACTGATTCA 1269
||||| ||| ||| | ||| ||| ||| |||
Db 40 AATCACTACTACTACAACACAACAACTAATCCAAACAGGTTCA 1

RESULT 14
US-08-145-705A-34/c
; Sequence 34, Application US/08145705A
; Patent No. 5489513
; GENERAL INFORMATION:
; APPLICANT: Springer, Wolfgang; PL

APPLICANT: L bberding, Antonius
 TITLE OF INVENTION: SPECIFIC GENE PROBES AND
 TITLE OF INVENTION: PROCESSES FOR THE DIAGNOSTIC
 TITLE OF INVENTION: INVESTIGATION OF CANDIDA
 TITLE OF INVENTION: ALBICANS
 NUMBER OF SEQUENCES: 44
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SPRUNG HORN KRAMER & WOODS
 STREET: 660 White Plains Road
 CITY: Tarrytown
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10591-5144

10591314
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.5 inch, 1.4 MB storage
 COMPUTER: NEC PowerMate 1 Plus
 OPERATING SYSTEM: DOS
 SOFTWARE: Wordperfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/145,705A
 FILING DATE: October 28, 1993

Query Match	2.2%	Score 82.4;	DB 1;	Length 100;
Best Local Similarity	89.0%;	Pred. No. 2.5e-11;		
Matches 89; Conservative	0;	Mismatches 11;	Indels 0;	Gaps 0;

QY 1030 TTACCATTC AATCCCAAGTGTGATAAAACCAAAACAATCG 1069
|||||
D_b 40 TTACCATTC AATCCCGATGTTGACAAAACATAAAACAATTG 1

RESULT 15
US-08-145-705A-33/C
; Sequence 33, Application US/08145705A
; Patent No. 5489513

STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.4 MB storage
COMPUTER: NEC PowerMate 1 Plus
OPERATING SYSTEM: DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/145,705A
FILING DATE: October 28, 1993
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: German P 42 36 708.5
FILING DATE: October 30, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 8885-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
TELEX:
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Candida albicans
US-08-145-705A-33

Query Match 2.0%; Score 76; DB 1; Length 100;
Best Local Similarity 85.0%; Pred. No. 9.3e-10;
Matches 85; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 1070 AAATTTTGACCTATTCCAACTACCATCACAACATCATATGTGGTGACTACTT 1129
DB 100 AAATTTTGACCCCATTCACAACTACCATCACAACATCATATGTGGTGACTACTT 41
QY 1130 CCTATCTGACTAAGACTGACCAATTTGGTGAACAGCTAC 1169
DB 40 CCTACAGACCCAAACTGTACCAATAGGACAAACTGCTAC 1

Search completed: June 20, 2003, 06:16:05
Job time : 134 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 05:18:05 ; Search time 588 Seconds
(without alignments)
9448.434 Million cell updates/sec

Title: US-09-715-876-7
Perfect score: 3786
Sequence: 1 atgtctcaacaattacatt.....tgctctgttcatttagta 3786

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues
Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	78.4	2.1	5403	10	US-09-745-008-33
2	74.4	2.0	684973	10	US-09-263-959-1
c 3	62.2	1.6	2014	10	US-09-842-552-22
4	55.4	1.5	436	10	US-09-864-761-2885
5	54.6	1.4	867	10	US-09-216-393-340
6	54.6	1.4	867	10	US-09-216-393-342
7	54.6	1.4	1397	10	US-09-216-393-343
c 8	54.6	1.4	1397	10	US-09-216-393-345
9	52.8	1.4	4104	10	US-09-801-368-107
c 10	52.2	1.4	1059	12	US-10-073-256-54
c 11	52	1.4	2015	10	US-09-842-552-79
12	50.8	1.3	15720	9	US-10-025-380-1058
13	50.8	1.3	15720	10	US-09-922-217-1058
14	50.8	1.3	15720	10	US-09-833-263-1058
15	50.6	1.3	1236	9	US-10-077-584-3
16	50.2	1.3	4197	10	US-09-137-531-7
17	50.2	1.3	4197	10	US-09-137-531-8
c 18	48.8	1.3	3183	10	US-09-955-909-1
c 19	48.6	1.3	2614	9	US-09-822-846-491

20	48.2	1.3	1101	10	US-09-874-062-2	Sequence 2, Appli
c 21	48.2	1.3	6604	10	US-09-880-107-1748	Sequence 1748, Ap
c 22	48	1.3	3331	10	US-09-864-761-19481	Sequence 19481, A
23	48	1.3	7104	10	US-09-815-242-4580	Sequence 4580, Ap
24	48	1.3	7107	10	US-09-815-242-8291	Sequence 8291, Ap
25	47.8	1.3	460	10	US-09-864-761-19383	Sequence 19383, A
26	47.8	1.3	2824	12	US-10-124-557-13	Sequence 13, Appl
27	47.8	1.3	3066	12	US-10-124-557-83	Sequence 83, Appl
28	47.8	1.3	3117	12	US-10-124-557-73	Sequence 73, Appl
29	47.8	1.3	3148	12	US-10-124-557-57	Sequence 57, Appl
30	47.8	1.3	3420	12	US-10-124-557-103	Sequence 103, App
31	47.8	1.3	3813	12	US-10-124-557-43	Sequence 43, Appl
32	47.8	1.3	3936	12	US-10-124-557-41	Sequence 41, Appl
33	47.8	1.3	3945	12	US-10-124-557-141	Sequence 141, App
34	47.8	1.3	3945	12	US-10-124-557-49	Sequence 49, Appl
35	47.8	1.3	3963	12	US-10-124-557-45	Sequence 45, Appl
36	47.8	1.3	3963	12	US-10-124-557-59	Sequence 59, Appl
37	47.8	1.3	4065	12	US-10-124-557-47	Sequence 47, Appl
38	47.8	1.3	4086	12	US-10-124-557-39	Sequence 39, Appl
39	47.8	1.3	4092	12	US-10-124-557-51	Sequence 51, Appl
40	47.8	1.3	4215	12	US-10-124-557-61	Sequence 61, Appl
41	47.8	1.3	4575	12	US-10-044-090-303	Sequence 303, App
42	47.8	1.3	5008	12	US-10-124-557-1	Sequence 1, Appli
c 43	47	1.2	5361	9	US-09-742-096-2	Sequence 2, Appli
c 44	47	1.2	6060	7	US-08-781-986A-534	Sequence 534, App
c 45	47	1.2	6152	9	US-09-742-096-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-745-008-33/c
; Sequence 33, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Marcio A.
; TITLE OF INVENTION: T. Cruz-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09745.008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 5403
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
US-09-745-008-33

Query Match		2.1%	Score 78.4;	DB 10;	Length 5403;
Best Local Similarity		42.2%	Pred. No. 3.8e-07;		
Matches 442;		Conservative	0;	Mismatches 606;	Indels 0; Gaps 0;
QY	1309	ACTGTTAGTACTACTGATATTTGGTCTCAGTCTCTTCTGCTACACCACTACAGTACTGCT	1368		
Db	3789	ACCGTGGGCACCTGCTGTCAGCGGGAGTCTGTCAGCGGCGTACTGTGGGCACCTGCTGTCAGCGGG	3730		
QY	1369	CTCTCCAGTGGTACCCGATATCTGTGATTTATCAGAGAGCCACCAACCATCTGTCACACT	1428		
Db	3729	AGTCGAGGGCGTACCGTGGGCACCTGCTGTCAGCGGGAGTCTGTCAGCGGCGTACTGTGGGCAC	3670		
QY	1429	ACTGAATATTTGGTACAAATCTTTTGGTACTACTACTGCTGCTCTCCAGTGGT	1488		
Db	3669	GCTGTCAACGGGAGTCGAGGCGCTACTGTGGGCACCTGCTGTCAGCGGGAGTCTGAGGCGT	3610		
QY	1489	ACTGACTACTGATTTATTCACAGACACCAACCAATCCAACTCTACTCAACCGAGTATTGG	1548		
Db	3609	ACCGTGGGCACCTGCTGTCAGCGGGAGTTGAGGCGTACTGTGGGCACCTGCTGTCAGCGGG	3550		

;	LENGTH: 2014	
;	TYPE: DNA	
;	ORGANISM: Bacillus anthracis	
US-09-842-552-22		
Query Match	1.68;	Score 62.2; DB 10; Length 2014;
Best Local Similarity	42.3%;	Pred. No. 0.00092;
Matches 402;	Conservative 0;	Mismatches 548; Indels 1; Gaps 1;
Qy	1354	ACTACAGTTACTGCTCTCTCCAGGTGGTACCGATACGTGTGATTATCAGAGAGCCACCAAAAC 1413
Db	1356	ACAACTGGTTGGTTCTTCTACAGCTGGTGGCTCTCGCAACTGGTGTGTTCTTCTACGATC 1297
Qy	1414	CATACTGTCACTACTACTGAATATTGGTCACAATCTCTTGCTACTACTACTACTGTGTACT 1473
Db	1296	GATGTTTCCNCTGCAACTGGTGTCTTCTACGACCGGATGTTTCTCTGCAACTGGTGTG 1237
Qy	1474	GTCCTCCAGGTGGTACTGACTCAGTAATATCAGAGAACCCAAATCCAATCTGCAC 1533
Db	1236	TCCTTGTGCAATCGGTGCTTCTTACAACTGATTGTCTTCTACGACCGGTGCTTCTTCT 1177
Qy	1534	ACAAACGAGTATTGGTCTCAATCCTTTGGCTACTACTACTACAGTACTGTGCTCTCCAGGT 1593
Db	1176	ACAATCGGTGTTTCTTCTACAACGTGATTGTTCTTCTACGACCGGTGCTTCTTCTGCACT 1117
Qy	1594	GGTACTGACATCAGTAATATCAGAGAACCTCCAAACCCAACTGTGCACCACTCTGAATAT 1653
Db	1116	GGTGTGTTCTTACAATCGGTGTTTCTTCTACAACGTATGTTCTTCTACGACCGGTGCT 1057
Qy	1654	TGPTCCCAATCTTACGCAACCACTACTGTGACTGCTCCTCCAGGAGCACTGACTCA 1713
Db	1056	TCCTCTGCAACTGGTGTCTTCTTCTACAATCGGTGTTCTTCTACAACGTATGTTCTCTCC 997
Qy	1714	GTAATATCAGAGAACCCAAACCCACAC-TGTCACTACTACTGAATACTGGTGCACATC 1772
Db	996	GCAATTAACACATCCGCAAGAGCTTTGAACATCTTTTTTCGTTGTCATCAGCGCACGCA 937
Qy	1773	ATAATGCCACCATACCACTCTGAATCTGCACCACAGGTGGTACTGACACTGTTATCATCTAG 1832
Db	936	ACATGATGAATTCATCTCTTTGTTGTTCAATGCTTCATTTAACACCATTTCTTGTCTATTC 877
Qy	1833	AGAGCCACCAACCACTGCTACTACTACTAGTATGTTGTCATCGTTTGTCTACTAC 1892
Db	876	CTCAAGTGTACAGGTGCCACTTCTTCTGCTGCTTCTGTTCTGCAATTACTTCTACTTTC 817
Qy	1893	CACAACTGTAACTGGTCCCAAGTGGCACTGATAGTATTATCATAGGGAACCCACAAA 1952
Db	816	TTCTGACTCTTCTGTTTCTGCAATTAGTCCACTTCTTCTTAATTTCTTCTAATTTTCTGT 757
Qy	1953	CCCAACTGTCCAGCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTC 2012
Db	756	TTCTGCAATTACTTCCACTCTTCTTAACTCTCTGTTTCTGCAACTACTTCCACTTCTTC 697
Qy	2013	CGTCTCCACCTGGTGAACACTGATACCGTCTTATCAGAGAGCCACCAACCATACTGTGAC 2072
Db	696	TAAATCTCTGTTTCTGCACTACTTCCACTTCTTCTTAACTCTCTGTTTCTGCAATTC 637
Qy	2073	TACTACTGAATAGTGGTCTCAATCAWATGTTACACCACTGTTTACTGCAACCACTGG 2132
Db	636	TTCTACTTCTTCTGACTCTTCTGTTTCTGCAATTACTTCTACTTCTTCTGACTCTTCTGT 577
Qy	2133	TGAACCGATACCGTCTTATCAGAGAGCCCAACCAACCATACTGTCACTACTACTGAATA 2192
Db	576	TTCCGCAATTACTTCTACTTCTTCTGACTCTTCTGTTTCTGCAATTACTTCTACTTCTTC 517
Qy	2193	CTGGTCTCAATATGCTACAACCACTGTTTACTGCAACCACTGGTGGTACCGATAC 2252
Db	516	TGACTCTCTCTGTTCTGCAATTACTTCTTCTTCTGACTCTTCTGCTTCTGCAATTC 457
Qy	2253	TGTTATCATTTAGAGAGCCCAAAATCCAACAGTTTACTTACTTACTGAATATG 2303
Db	456	TTCTACTTCTTCTGACTCTTCTGTTTCTGCAATTACTTCTCCACTTCTTCTCTCG 406

Qy	2073	TACTACTGAATACTGGTCTCAATCATATGCTACAACCAACCACTGTTACTGCACCACTCG	2132
Db	636	TTCTACTCTTCTGACACTCTTCTGTTCTGCAATTACTTCTACTTCTTCTGACACTTCTGT	577
Qy	2133	TGAACCGATACCGTTCCTTATCATGAGAGCCACCAACCATACTGTCTACTACTTGAATA	2192
Db	576	TTCCGCAATTACTTCTACTTCTTCTGACTCTCTGTGTTCTGCAATTACTTCTACTTCTTC	517
Qy	2193	CTGGTCTCAATATATGCTACAACCACTGTTTACTGCAACCACTGGTGGTACCGATAC	2252
Db	516	TGACTCTCTGTTTCTGCAATTACTTCTACTTCTCTGACTCTTCTGTTTCTGCAATTAC	457
Qy	2253	TGTTATCATTTAGAGAGCCCAATCCAAGTTTACTACTTGAATATTG	2303
Db	456	TTCTACTTCTCTGACTCTTCTGTTTCTGCAATTACTTCCACTTCTTCTCTGG	406

Matches	152;	Conservative	0;	Mismatches	161;	Indels	0;	Gaps	0;
Qy	965	CTAACGGTATTGTCAATTTGGTCTACAACTAGAACAGTTACAGACAGTACCACTGCTGTGCA	1024						
Db	68	CTATTGTCACTACTATTACCACCTAGCACTAGTACTACCACTATCATCACTACCATCA	127						
Qy	1025	CTACTTTACCAATTCATCCAAAGTGTGATATAAACCAACAAATCGAAATTTTGCACACCTA	1084						
Db	128	CTACTAATATCATCTATTACCACCTACTATTACCATTACCAAGGACTATCATTTACTACTACTA	187						
Qy	1085	TTCCAAACCACTACCATCACTCAATTCATATCTTGGTGTGACTACTTCTATCTGACTAAGA	1144						
Db	188	CCAGCTCCACCACTTACTACTACTAATGCTATCATTTACTACTTCTTATTAATTA	247						
Qy	1145	CTGCACCAATTTGGTGAACACAGCTACTGTTATTGTTGATGTGCCATATCATCTACTACCACAA	1204						
Db	248	CTCCTACTCCTACCACTTACTACTACTAGTAGTAGTATCACCATTACTACTACTATCATTA	307						
Qy	1205	CTGTTACCAAGTGAATGGACAGGAGCAATCTACTACCACCACAACTCGTACCAATCCAACTG	1264						
Db	308	CTATTACCACCACTTACTACTACTATCATCTACTACTACCACCACTTACTACTACTG	367						
Qy	1265	ATTCAATTGACAC	1277						
Db	368	CTGTCACTACTAC	380						
RESULT 5									
US-09-216-393-340									
; Sequence 340, Application US/09216393									
; Patent No. US2001001447A1									
; GENERAL INFORMATION:									
; APPLICANT: Milhausen, Michael James									
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND									
; TITLE OF INVENTION: USES THEREOF									
; FILE REFERENCE: TX-1-C2									
; CURRENT APPLICATION NUMBER: US/09/216,393									
; CURRENT FILING DATE: 1998-12-18									
; EARLIER APPLICATION NUMBER: 08/994,825									
; EARLIER FILING DATE: 1997-12-19									
; NUMBER OF SEQ ID NOS: 364									
; SOFTWARE: PatentIn ver. 2.0									
; SEQ ID NO 340									
; LENGTH: 867									
; TYPE: DNA									
; ORGANISM: Toxoplasma gondii									
; FEATURE:									
; NAME/KEY: CDS									
; LOCATION: (1)..(867)									
US-09-216-393-340									
Query Match 1.4%; Score 54.6; DB 10; Length 867;									
Best Local Similarity 45.5%; Pred. No. 0.028;									
Matches 195; Conservative .0; Mismatches 234; Indels 0; Gaps 0;									
Qy	3017	CTGATTTTGACTAGTGAACCAACAAATACCAAGAGAACCACTACTATTATCACTACTT	3076						
Db	395	CTGGTTTCAATAGGCCAAGGAGGTCAAGAGAGGGAAAGACTACAACTTACAAACGTT	454						
Qy	3077	CAAACTCCATCTAGTGAAGATATACCACTACTCTCAACCTTACAGGTGATATGGAGACAATA	3136						
Db	455	CATCCACAAGTACGAGTACACGACGACCACTACTACTACTTACCCTTACCACCA	514						
Qy	3137	CTTCATCAACCAATCCAGTTCCAACTGTGCGACACAAGTACTTTAGCATCTCGAAGTGAAG	3196						
Db	515	CTACGACTACTACTACAACTACGACACCAACAACTACCAACCAACCAACCACTACAC	574						
Qy	3197	AAGACAACAAAAGCGTTCTCATGATFCAGCATCCACAAGTTTGAACCAAGTATGGTG	3256						
Db	575	CACAAACACGACAAACCAACCAACAACTACACCAACCAACGACACCAACCAACCAACAA	634						
Qy	3257	AAAATCTGGATTAACTACTTCTACTGAAATTTGAAGCTTACAAACCAACCAAGTCCTACAGAAG	3316						

Search completed: June 20, 2003, 08:40:08
Job time : 595 secs

RESULT 15
US-10-077-584-3
; Sequence 3, Application US/10077584
; Publication No. US20030073610A1
; GENERAL INFORMATION:
; APPLICANT: LINDQUIST, SUSAN
; APPLICANT: KROBITZSCH, SYLVIA
; APPLICANT: OUTEIRO, TIAGO F.
; TITLE OF INVENTION: YEAST SCREENS FOR THE
; FILE REFERENCE: ARCD:3670US
; CURRENT APPLICATION NUMBER: US/10/077,584
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269,157
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:11:32 ; Search time 47 Seconds
(without alignments)
3572.251 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQOFTLLFLVLSIASAKTI.....SLIQHSTWLYGLITLTLSLFI 1260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	938	14.4	1537	15	Yeast 4.7 kb agglu
2	721.5	11.1	5179	22	C899P predicted am
3	685	10.5	2586	22	Drosophila melanog
4	651	10.0	2137	23	Staphylococcus epi
5	640.5	9.9	2344	22	Staphylococcus aur
6	601	9.3	1795	22	AAU37120
7	597	9.2	1296	23	Drosophila melanog
8	597	9.2	1296	23	Human novel polype
9	553.5	8.5	2870	21	Human novel polype
10	553.5	8.5	3178	21	Caenorhabditis ele

11	545.5	8.4	957	21	AAV59288	Human MUC11 polype
12	545.5	8.4	957	22	AAW24513	C900P predicted am
13	534	8.2	849	17	AAW06725	FLO1 protein, invo
14	516	7.9	894	15	AAW47578	Flocculation prote
15	516	7.9	894	15	AAW58754	S. cerevisiae FLO1
16	507	7.8	1532	21	AAW40945	Human ORFX ORF709
17	488	7.5	1721	19	AAW48299	Cryptosporidium pa
18	488	7.5	1721	21	AAW11727	Portion of Cryptos
19	488	7.5	1721	23	ABJ04045	C parvum GP900 pro
20	487	7.5	1837	21	ABJ11726	Cryptosporidium pa
21	487	7.5	1837	23	ABJ04044	C parvum GP900 pro
22	468.5	7.2	862	15	AAW60563	yeast 2.6 kb agglu
23	463	7.1	688	22	ABB30137	Peptide #2788 enco
24	463	7.1	688	22	ABB35307	Peptide #2813 enco
25	463	7.1	688	22	ABB20749	protein #2748 enco
26	463	7.1	688	22	AAW56138	Human brain expres
27	463	7.1	688	22	AAW68511	Human bone marrow
28	463	7.1	688	22	AAW16315	Peptide #2749 enco
29	463	7.1	688	22	AAW28810	Peptide #2847 enco
30	463	7.1	688	22	AAW04053	Peptide #2735 enco
31	463	7.1	688	23	ABG38092	Human peptidase enco
32	450.5	6.9	498	22	ABW58595	Drosophila melanog
33	435.5	6.7	1045	22	ABG08332	Novel human diagno
34	428	6.6	1237	21	AAW81609	Streptococcus pneu
35	417.5	6.4	746	22	ABW59201	Drosophila melanog
36	417	6.4	560	22	AAU37464	Staphylococcus aur
37	404.5	6.2	2035	15	AAW57141	Host cell factor p
38	404	6.2	2570	22	ABG06375	Novel human diagno
39	402	6.2	2112	22	ABW60403	Drosophila melanog
40	392	6.0	502	22	AAU34408	Staphylococcus aur
41	392	6.0	1638	20	AAW00138	Enterococcus faeca
42	392	6.0	1638	20	AAW00140	Enterococcus faeca
43	392	6.0	1638	20	AAW00142	Enterococcus faeca
44	392	6.0	1638	23	ABP43357	Enterococcus faeca
45	392	6.0	1638	23	ABP43359	E faecalis EF068 p

ALIGNMENTS

RESULT 1
AAW60562
ID AAW60562 standard; Protein; 1537 AA.
XX
AC AAW60562;
XX
DT 13-APR-1995 (first entry)
XX
DE Yeast 4.7 kb agglutination gene FLO1L.
XX
KW Yeast; agglutination; FLO1L.
XX
OS Saccharomyces cerevisiae ABXL-1D.
XX
PN WO9419475-A.
XX
PD 01-SEP-1994.
XX
PF 24-FEB-1994; 94WO-JP00290.
XX
PR 26-FEB-1993; 93JP-0038871.
XX
PA (PANT-) PANIMOLABORATORIO BRYGGERILABORATORIUM.
(SAPB) SAPPORO BREWERIES.
XX
PA Keraenen S, Ogawa M, Onnela M, Penttila M, Takata Y;
PI Watarai J;
XX
DR WPI: 1994-294338/36.
DR N-PSDB; AAQ71390.
XX
PT New yeast agglutination genes and yeast contg. them - impart
agglutination properties to facilitate removal from fermentation

pt media

Disclosure; Page 43-48; 75pp; English.

The agglutination gene is called FLO1. Saccharomyces cerevisiae includes an agglutination gene of 4.7 kb (FLO1L) and an agglutination gene of 2.6 kb (FLO1S). FLO1L is the intact FLO1 gene on chromosome 1, and FLO1S is the FLO1L gene with a portion of the ORF deleted in frame. FLO1L imparts a relatively strong agglutinative property to the host yeast into which it is introduced, while FLO1S imparts a weaker agglutinative property.

Sequence 1537 AA;

Query Match 14.4%; Score 938; DB 15; Length 1537;
 Best Local Similarity 26.0%; Pred. No. 3 6e-40;
 Matches 417; Conservative 244; Mismatches 523; Indels 418; Gaps 85;

2 LOQFTLLFYLSTASAKT-----ITGVDFDSF-----NSLTWSNAANYAF-----40
 10 LAVFTLLAL-TSVASGATEACLPAGQKSGMNIIFYQYSLKDSSTYSNAYMAYGYASKT 68
 41 -----KGGCYPW-----NAVLGWSLD--G 58
 69 KLGSGVGQTDISIDYNIPCVSSSGTTPCQEDSYGNWCKGMGACNSQGIAYWSTDLEF 128
 59 TSANPGDTFTLNN-----PCVFKYTTSQ-----TSVD 85
 129 FYTTPNV-TLEMTGYFLPQGTGYTFKATVDDSDAILS VGGTAFNCCAAQQPPTSIN 187
 86 LRADGVK-----YACQFYSGEETFTSTLCTVNDALKSIKAFGT-----VTLP 131
 188 FTIDGKPKWGLSPNPNIEGTVMYAGYYPM-----KVYSSNAVSWGLTPISVTLTP 238
 132 IAFNVGSGSTDLDSKCFAGTNTVTFNDGDKDISIDVEKSTV-DPSAYLYASRYM 190
 239 -----DCTTVSDDFG-----YVYSEDD-----DLSSNCTVPPDSNYA-VSTTT 277
 191 PSLNKVTLTLFVAPQCENGYSCTGMPSSNGSDVAIDCSNIHIGITKGLNDWNPVSSSEF 250
 278 TTTEPWGTFTSTSTMTVGTGNGVPTDETVIVIRTPPTASTIITTEPNWSTFTSTST 337
 251 SYTKTCTSGIKYQNVAGRPFDAYISATDVNQYLYATNDYTCAGSLQSKPFTL 310
 338 ELTVVGTGNGVRDEII-----VIRTPPTATTITTEPNWSTFTSTSTEL-----384
 311 RWIGYKNSDAGSNGI--VIVATRTVDTSTAVTTL-PFNPSVDKTKT-----355
 385 -----TTVTGNGLPTDETVIVIRTPPTATTITTEPNWSTFTSTSTMTVVTGNTGL 438
 356 -----IEILOPIPTTTIT-----SYVGVTTSTYLTAKTA-----PIGETATVIVDVPYH 398
 439 PTDETVIVIRTPPTATTITTEPNWSTFTSTSTMTVVTGNGLPTDET-IIIVIRTPIT 497
 399 TTT--TIVTSEWGTIT-----TTTTRTN-PTDSIDTVVQVPLPNPT-VSTTEYWSQS 447
 498 ATTAMTTQPNWDTFTSTSTMTVGTGNGLPTDE-TIIVIRTPPTATTITTEPNWSTFTSTSTMT 556
 448 FATT-----TTVTAPPG-GTD--TVIIRPPNHT--VTTTEYWSQSFAIT-----TTVTAPP 494
 557 FTSSTSTMTVVTGNGLPTDETVIVIRTPPTATTITTEPNWSTFTSTSTMTVVTGNTGN 616
 495 G-GTDS--VVIIRPPNPT--VTTTEYWSQSFAIT-----TTVTAPPG-GTDS--VVIIRPP 542
 617 GLPTDETVIVIRTPPTATTITTEPNWSTFTSTSTMTVGTGNGLPTDET-IIIVIRTPIT 676
 543 NPT--VTTTEYWSQSFAIT-----TTVTAPPG-GTDS--VVIIRPPNHT--VTTTEYWSQS 591
 677 TATTAMTTQPNWDTFTSTSTMTVGTGNGLPTDETVIVIRTPPTATTITTEPNWSTFTSTSTMT 736
 592 YATT-----TTVTAPPG-GTD--TVIIRPPNHT--VTTTEYWSQSFAIT-----TTVT-- 635
 737 FTSSTSTMTVVTGNGLPTDETVIVIRTPPTSEGLISTTTTEPNWSTFTSTSTMTVVTGNTGN 796

QY 636 GPPSGTDVVIIRPPNP--TVTTEYWSQSFAIT-----TTITAP-----PGETDVLIREPP 686
 D 637 GPPSGTDVVIIRPPNP--TVTTEYWSQSFAIT-----TTITAP-----PGETDVLIREPP 686
 D 638 GPPSGTDVVIIRPPNP--TVTTEYWSQSFAIT-----TTITAP-----PGETDVLIREPP 686
 QY 687 NHTV--TTEYWSQSFAIT-----TTVTAPPG-GTD--TVIIRPPNHT--VTTTEYWSQS 735
 D 688 NHTV--TTEYWSQSFAIT-----TTVTAPPG-GTD--TVIIRPPNHT--VTTTEYWSQS 735
 D 857 SEGLISTTTTEPNWSTFTSTMTVGTGNGLPTDETVIVIRTPPTSEGLISTTTTEPNWSTFTSTMT 916
 QY 736 YATT-----TTVTAPPG-GTD--TVIIRPPNHT--VTTTEYWSQSFAIT-----TTVTAPP 785
 D 917 FTSSTSTMTVVTGNGLPTDETVIVIRTPPTSEGLISTTTTEPNWSTFTSTMTVVTGNTGN 976
 QY 786 -----DVIIRPPNHT--VTTTEYWSQSFAIT-----TTITAP-----PGETDVLIREPP 686
 D 977 GPPSGTDVVIIRPPNP--TVTTEYWSQSFAIT-----TTITAP-----PGETDVLIREPP 686
 QY 817 --HYNSTSD--LSTFESSMNTPTSISSDGLMLLSTTLVTESETT-----TELICSDGRE 869
 D 1037 SEGLVTTTTEPNWSTFTSTMTVGTGNGLPTDETVIVIRTPPTAIISSSLSSSSSGOI 1096
 QY 870 CSRLLSSSGIVTNP--DSNESSIVTSTVPTASTMSDLSLSTDCISATSDNVSKSGSVT 927
 D 1097 TSITSSSRPIIT-PFYPSNGTSSVSSSVTSSSLFTSSPVISSSVSSSTTTSTSTIF 1155
 QY 928 TETSVTIQTTPNPLSSSVTSLTQSSIPSVSESKVTFTSNGDNQSGTHDSQSSTEI 987
 D 1156 SESSKSSV--IPTSSSTSGSESETSAGSVSS-----SFISSESKSTYSSSS---L 1205
 QY 988 EIVTSSST-----KVLPPVSSNNDLTSEPTNTRPOTTLSTTS-----NSITEDIIT 1035
 D 1206 PLVTSATTSQETASSLPPATT-----TKTSEQTLVTVTSCESHVCTESISPAIVS 1256
 QY 1036 SQPTGNGNTSTN--PVPTVATSTLASASEDNKSGSHASASTSLKPSMGNSGLTSS 1093
 D 1257 TATVTVSGVTETVWCPISITETTKQTGTEQTTETTKQTIVTVTSSCESDVCCKTAS 1316
 QY 1094 TEIATETPTTEAPSPAVSSGTDVTEPTDTEPTTLSTT-----SKTINSELVAT-- 1145
 D 1317 PAIVTSSTATINGVTETWCPISIT--TESRQTLVTVTSCESGVCSETASPAIVSTA 1374
 QY 1146 -----QATNENGKSPDLDLSSLTGTP--SASTSANSSELVTSVGGGA 1188
 D 1375 TATVNDVTVPTWRPQTANEESSVSKMNSATGETTTNTLAAETTTNTVAAETITNTGAA 1434
 QY 1189 -----VASANDQSHSTSV-----TNSNIVSNTPTTLLSQQVTS 1224
 D 1435 ETKVTVTSSLSRSHAEOTASATDVIGHSSSVSVSETGNTKSLTSSGLSTMSQOPRST 1494
 QY 1225 SP-----STNTFIASYDGSIIQHSWLYGLITLLSLFI 1260
 D 1495 PASSVMGVSTASILEISTYAGSANSLLAGSLSVFIASLLAI 1536

RESULT 2
 AAM24516
 ID AAM24516 standard; Protein; 5179 AA.
 XX AAM24516;
 AC AAM24516;
 DX 12-OCT-2001 (first entry)
 XX C899P predicted amino acid sequence.
 XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 XX immunogenic; gene therapy; vaccine; colonic cancer.
 XX Homo sapiens.
 XX WO200149716-A2.
 PN 12-JUL-2001.
 PD 12-JUL-2001.
 XX

QY 1072 --SHEAS-----TSKPSMGNSGLT-----TS 1093
 Db 1400 DISQSSSNVLDNNYNDLFRWANNPPYLTPLWKSVEQMAITAPLSLRPPQOQTS 1459
 QY 1094 TEIANTTSPEAPSVSSGTDVTEPTDREQPTTLSTSTKTNSE-----LVA 1143
 Db 1460 SD-EGQTTTS-SSAPVVDISQSSSSNGDGNST-QSSTTTTNTTSSDGGESTLLDPVVE 1516
 QY 1144 TQATNENGKSPSDDLSSLTGTA---STANSSELV-----TSGSVT 1185
 Db 1517 VSQGTN---GDNNTQSSSTTTTSSDEGQTTSSSDPVEVAQSSNGDGNSTQSSST 1573
 QY 1186 GCAVASANDQSHST-----SVTNSNSIVSNTPTQTLSSQOVTSSSPST 1228
 Db 1574 TTTTSSDGGESTSSDPVVEVSQGTNGDGNSSSTQSSSTTTTSSDEGQTTSSSDPVP 1633
 QY 1229 NTFIASTYDGSIIQHSST 1247
 Db 1634 EVAQSSSSNGDGNSTQSSST 1652

RESULT 4

ABP39618
 ID ABP39618 standard; Protein; 2137 AA.
 AC ABP39618;
 XX
 XX
 DT 24-JUL-2002 (first entry)
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4463.
 XX
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KW antibacterial; gene therapy.
 XX
 XX Staphylococcus epidermidis.
 OS
 PN US6380370-B1.
 XX
 PD 30-APR-2002.
 XX
 PF 13-AUG-1998; 98US-0134001.
 XX
 PR 14-AUG-1997; 97US-055779P.
 PR 08-NOV-1997; 97US-064964P.
 XX
 XX (GENO-) GENOME THERAPEUTICS CORP.
 PA

PI Doucette-Stamm LA, Bush D;
 XX
 XX WPI; 2002-381255/41.
 DR N-PSDB; ABN92163.
 XX
 XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 XX
 XX Disclosure; SEQ ID 4463; 267pp; English.
 PS
 XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life
 CC cycle or inhibit S. epidermidis infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.
 XX
 SQ Sequence 2137 AA;
 Query Match 10.0%; Score 651; DB 23; Length 2137;

Best Local Similarity 24.2%; Pred. No. 3.7e-25;
 Matches 314; Conservative 252; Mismatches 541; Indels 192; Gaps 37;
 QY 19 TITGV-----FD-SNSLTWNAANYAFKPGPYTWNVLGSLDGTSANPGDTFTLNP 72
 Db 745 TTTGLPQGLKFFDASTNSIV-----GTPTQIGTNTTITESTDASGNK----- 785
 QY 73 CVFKYTTSTSVDLTADGVKATCOFYSGEEFTTSTLCTVNDALKSIKAFGTVLP 132
 Db 786 -----TTTKINYEVRNSASDSTSTSVNSVSTISNST-----SLSDSVKASQSL---- 831
 QY 133 AFNVGSGSTSDLEDKSCFTAGTNTVFNDDKDISIDVEFEK--STVDPSAYLAYSRVM 190
 Db 832 -----STKSLSSESLSASTSNSTSIQASESASTSKQLSESASTSDSASESARKS 882
 QY 191 PSLNKVTTLEVAPOQENGYSCTMGCFSSNGDVAIDCSNIHICITKGLNDWNPVSESEF 250
 Db 883 ESTSKSTSL-----SESTSTSVSDSASVSTSESASTSTSVSGSTSTSID----- 927
 QY 251 SYTKTCTSGIQTQYONVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSLQSKPFT- 309
 Db 928 -STSTSDSASIKASESASTSKLLSESVSTSDSASTSTSDSASTSTSLSKSTST 986
 QY 310 -----LRWTGYKNSDAGSNGIVIVATRTVTDSTTAVTTLPFNPSVDKTKTI 356
 Db 987 VSDSTSTSDSASTSTSESDSASTSL-SESTSTSVSDSTSTSTSDSASTSTSDSASTSESEN 1045
 QY 357 EILQPIPTTITTSYGVTTSYLTKTAPIGETATVIVDPYHTTHTTIVTSEWGTIITTT 416
 Db 1046 SKSTSLSEST-STSLSGSTASSTSDSA-----STSTSESDSTSTSTSLSESTSLSGST 1099
 QY 417 RNPNTDSIDTVVQVPLPNTVTTEYWSQSFATTTTVPAPGGTDTVIREPPNHTVTT 476
 Db 1100 SASTSDAST-----STSESDSTSESTSLSESLSTSVSDSTSTSESTSESTST 1147
 QY 477 TEYWSQSFATTTTVPAPGGTDSVIREPPNPTVTTEYWSQSFATTTTVPAPGGTDSV 536
 Db 1148 SESENSASTSLSGSLSTSTSDSTSTSTSDSASTSTSESDSTSTSTSLSESTSLSDST 1207
 QY 537 IIREPPNPTVTTEYWSQSFATTTTVPAPGGTDSVIREPPNHTTHTTIVTSEWGTIITTT 596
 Db 1208 STSTSESDSTSESDSTSESTSTSVSDSTSTSDSASTSTSVSDSASTSTSVSDSASTSI 1267
 QY 597 TVTAPPGGDTVIREPPNHTTTEYWSQSFATTTTVPAPGGTDTVIREPPNPTVTTT 656
 Db 1268 SESLSTSVSDSTSTSDSASTSTSESDSTSTSESTSLSESLSTSVSDSTSTSDSASTST 1327
 QY 657 TEYWSQSFATTTTITAPPGTDTVIREPPNHTTTEYWSQSFATTTT-----VTAP 709
 Db 1328 SE--SESDASTSL--GSTSTSL-----SDSTSTST--SDSASTSTSESDSASTSL 1375
 QY 710 PGCTDTVIREPPNHTTTEYWSQSFATTTTVPAPGGTDTVIREPPNPTVTTEYWS 769
 Db 1376 SGSTSTSL-----SDSTSTST--SDSASTSTSVSDSASTSL-----SGSLSTSVSDS 1422
 QY 770 QGFATTTTVPAPGGTDTVIREPMSSSKISTSSNDITSIIPFSRPHVYNSTSDSLSTF 829
 Db 1423 TSTSTSDSASTSTSESDSASTSLSGS--TSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1474
 QY 830 ESSSMNTPTTSSDGMLLSTTLVTESETTTELLICSDGKESCLSSSSGIVTNPDSNESS 889
 Db 1475 ESNSTSTSTSESLSTSVSDSTSTSTSDSASTSTSVSDSASTSTSTSTSTSTSTSTSTST 1532
 QY 890 IVTS-TVPTASTMSDLSLSTDGIATSSDNVSKS-GVSVTTETSVTTIOTTPN--PLSSS 945
 Db 1533 TSTSDASTSTSVSESNSTSTSLSGSTSTSTSVSDSTSTSTSDSASTSTSESDSASTSTSS 1592
 QY 946 VTSLSLQSLIPSVSESESKVTTNSGCD-NQSGTHDSQSTSTETETVTTSTKVLPPVSS 1004
 Db 1593 ESVSTSVSDSTSTSESDSASTSTSVSDSASTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1652
 QY 1005 NFDLTASEPTNREQPTTLST 1064

Db 1653 ESDSDASTSLD-----STSTSVSESTSTSTSVASNSNSTSTSLSDSRSTSLSDSTS 1706
QY 1065 EEDNKSGS-----HESASTSLKPSNGE--NSGLTSTSTRIEATT-----TSPT 1104
Db 1707 TSTSESGSTSTSESDASTSLSESTSTISDSTSTSTSDASTSMVSDSNRSTSL 1766
QY 1105 EAPSPAVSGSTDTYTE---PTDTPREOPTTLSTSTKTNSELVATQ-----ATN 1149
Db 1767 DSTSTSVSDSTASTSEASTSTRESEASTSLSESTSTSVSDSTSTSTSDASTSTS 1826
QY 1150 ENGKGPSTDLTSLTGTSTASNSANSELVTSGVTGGAVASANDQSHSTSTVNSIV 1209
Db 1827 ESDNSESTSLSESTSTVSDSTSTAS-----TSASTSTSVSDSNASTSLSGSTSTSV 1882
QY 1210 SNTPQTLSQVTSSTSTNTFTASTYDGGSGSI-IQHST 1247
Db 1883 DSTSTSTASASTSTSESDASTSLSGSTSTSTSDST 1921

RESULT 5
AAU37120
ID AAU37120 standard; Protein; 2344 AA.
XX
AC AAU37120;
XX
DT 14-FEB-2002 (first entry)
XX
STaphylococcus aureus cellular proliferation protein #1290.

XX Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX

OS Staphylococcus aureus.

XX W0200170955-A2.

PN 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-6111495/70.

DR N-PSDB; AAS44979.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 12713; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The

CC invention is also useful for the identification of potential new targets

CC for antibiotic development. The antisense nucleic acids can also be used

CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.

CC The proteins can be used to screen compounds in rational drug discovery

CC programmes. The antisense nucleic acid sequence is also useful to screen

CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2344 AA;

Query Match 9.9%; Score 640.5; DB 22; Length 2344;
Best Local Similarity 22.2%; Pred. No. 1.4e-24;
Matches 323; Conservative 269; Mismatches 542; Indels 323; Gaps 46;

QY 17 AKTITGVFDSFNSLTWSNAANYAFKPGYPTWNAVGLWSLDGTSANPGDTFLNMPGVFK 76
Db 297 AITLGRIDSNSKFHFGKVNKNGYEGHGGGIGFAP-----SPG----- 339
QY 77 YTTTSQTSVDLTADGVKYATCQFYSGBEFTTFTLTCTVNDALKSSIK-----AFGTV 128
Db 340 -VLGETGLNGAAVIGGLSNAF--GFKLDYHN-TSTPNSSAKAKADPSNVAGGAGAF 395
QY 129 TLPIAFNVGGTSGSTOLED--SKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPDSL 186
Db 396 VTDSYGVASTVTSSTADNAAKLVQPTNN-TFQD-----FDINYNGDT----- 439
QY 187 SRVMPSLNKVTTLFVAPO-----CENGVTSGTMGFSFSSNGDVAIDCSNIHIGITK 236
Db 440 -----KMTVKYAGQWTRNIDWIAKSGTTFNLSMTASTGG-ATNLQVQFG--- 487
QY 237 GLNDWNPVSSSEFSYTKTCTSGNGIQIKYQNVAG-----YRPFIDAYISATDVNOY 288
Db 488 -----TFEYTESAVT--QVRVYDVTGKDIIPPKYSGNVQDVVIDNQSA 532
QY 289 TLATYNDYTCAGSRLOS-----KPFILRWT-----GKNSDAGS-- 322
Db 533 LTAAGNYTTSVDSSYASTYNDTNKTKMTNAGOSVYTFYTDVKAPTVTYVGNQTI 592
QY 323 NGIVIVATTRTVTDSTTAVTLPFPNPSVDKTKTIEILOPI-----PT 364
Db 593 NPVLTTDNGTGTNTVNTVGLPSGLSYDSATNSIGTPTKIQSVTVVVDQANNKST 652
QY 365 TTITTSYGVGTTSLTKTAPIGETATVI----- 392
Db 653 TTFTINWDTAPTATV---PIGDKSEVFSPISPIATQDNSGNAVTNTVTGLPSGLTF 709
QY 393 -----VDVPYHTTTTSEWTCITTTT-----TRNPTDSIDTVVQVPLP 436
Db 710 DSTNNTISGTPTNIGTSTITIVSDASGNKTTTTFKYEIVTRNSMSDSVST--SGSTQ 767
QY 437 TVSTTEYWSQSATT-----TAVTAPPGTDTVIIRE--PPNHTVTTEYWSQSATT 488
Db 768 SVSTSKADQSASTSTSGSINTSTASTSKTSVSLSDSVASAKSLSTSESNVSSTST 827
QY 489 TVTAPPGTDSVLIIREPPNPTVTTEYWSQSATTTAVTAPPGTDSVLIIREPPNPTVT 548
Db 828 SLVNSQSVSS--MSGSVSKTSLSDFISNSSSTKESVSTSTSDSLRSTSLSDSVSM 885
QY 549 TEYWSQSATTTVTAPPGTDSVLIIREPPNHTVTTEYWSQSATTTVTAP----- 601
Db 886 STSGLSKQSLSSTSTSDASTSQSVSDSTSIESTSELSSESGSTSEISISINSV 945
QY 602 -----PGTDTVIIREPPNHTVTTEYWSQSATTTAVTAPPGTDTVIIREPPNPTV 654
Db 946 SASTSKLESQSSTISLSTSDSKSMSTSELSSTSTSDSVSGSLVAGSQSVSTSTSDSM 1005
QY 655 TTEYWSQSATTTTITAPPGTDTVLIIREPPNHTVTTEYWSQSATT-----TTVT 707
Db 1006 STSEMISDMSTSGSLAASDKSMVSVMSTSQSGSTSELSDSISTSDSDSKSLST 1065
QY 708 APGETDTVL-----IREPPNHTVTTEYWSQSATTTVTAPPGTDTVIIREPPNPTV 762
Db 1066 SOSGSTSTSTSSSVRMSESQSTSGMSTSDSTSTSTSTSTSTSTSTSTSTSTSTSTST 1119

Db 798 TRKVAITTOKETPTSTTTITFRKTTNNPEPTSTKPTSTTP---KPTSTTPK 851
QY 826 LSTFESSMNTPTSISSDGLMLSTTLVTESEETTELICSDGKESRLSSSSGIVNPDOS 885
Db 852 TSVASST--EKTIISSP--KPTTEKSTENPTNSV-----KTSALTSSQRAVSTTS 900
QY 886 NESSI---VTSTVPTASTMSDLSLSDGISATSSDNVSKSGSVVTPPE-----TSVTTI-- 935
Db 901 EPTKTTQNIITTPKPTTLKTSQE---ATSTQKVSIT--VTIITTKATESPLTLST 954
QY 936 ---OTTPNPL-----SSSVTSLSLTSSIPSVSESKVFTFNGDNQSGTHDSQST--- 983
Db 955 EEPNTTPKPLRTTPTTTSVATRIIT--TTISESSTSTSTQPKSTTPTSTRTPKV 1013
QY 984 -----STEIEIVTSKVLPPVVSNTDLTSEPTNTREOPT--TLSTTSNSTEDITWS 1036
Db 1014 TTVIVSTQNTPTTTSKTSVITTPNPSPSTQRTTPTTTPROPTISITASTTSGTRIPPTT 1073
QY 1037 QP-----TGNDGNTSNTNPVPTVATSTLASASEEDNKSGSHE 1074
Db 1074 NPQNSTSDTLATVTRPPCPDPDSTSDKNTNACTQELQOV--NLLELQSPQKQEQFTH 1131
QY 1075 SASTSLKPSGENSG-----LNTSTEIEATTSPTREAPSPAVSSGTD-----V 1117
Db 1132 RHTALTGRNTLGGQEVDPYMDADPSSAEESGQATTAKATMTSLAAHLLQKLFHII 1191
QY 1118 TTEPTDREOPTTLSTTSKTNSEL-----VATTO-----ATNE 1150
Db 1192 STTPPSREHAPTPSPSSQSSORSRGVIAQMARHNLATSKPFIHSLRLSLOQLASTQ 1251
QY 1151 NGKSPSTDLTSLTGTSTASTSANSSELVTSVGTGAVASANDQSHSTVNSNSIYS 1210
Db 1252 KRSIPPKTLVTHNTKEPEDSEYDSE--TSEQYTDDEVDKTKTPRAMSSITVAALVP 1309
QY 1211 NTPQTTLSQ--QVTSSSPSTNFIAT-----YDGS GS 1241
Db 1310 AVPSSTTTEREPQKTSPPSPKATSKTSSTTTQPIETTTGDLLEYDSSGS 1355

RESULT 7

ID ABG66702 standard; Protein; 1296 AA.

AC ABG66702;

XX 30-AUG-2002 (first entry)

DT Human novel polypeptide #37.

DE

XX Human; inflammatory condition; shock; sepsis; immune response;
KW cancer; wound healing; central nervous system disease; haematopoiesis;
KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW bone degenerative disorder; periodontal disease; reperfusion injury;
KW lung fibrosis; liver fibrosis; autoimmune disease; bacterial infection;
KW allergic condition; thrombolysis; thrombosis; coagulation disorder;
KW fungal infection.

OS Homo sapiens.

PN WO200244340-A2.

XX 06-JUN-2002.

XX 30-NOV-2001; 2001WO-US47004.

XX 30-NOV-2000; 2000US-0028952.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;

PI Yamazaki V, Ujwal ML, Drmanac RT;
XX WPI: 2002-508509/54.
DR N-PSDB; ABK94926.
XX Novel nucleic acids and polypeptides for diagnosis, treatment of
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT disorders, cancer and promoting wound healing -
XX Claim 10; Page 604-607; 672pp; English.
XX The invention relates to human novel polynucleotides and associated
CC polypeptides. The polynucleotides and polypeptides are useful for
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC and cancer and for promoting wound healing. The sequences are used to
CC induce the proliferation of neural cells and regeneration of nerve and
CC brain tissue, and are useful for the treatment of central and peripheral
CC nervous system diseases and neuropathies, such as Alzheimer's disease,
CC Parkinson's disease, Huntington's disease and amyotrophic lateral
CC sclerosis. The sequences are involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC cell disorders and platelet disorders such as thrombocytopenia, tissue
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC disease. The sequences of the invention are also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial and fungal
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC and coagulation disorders. Sequences ABG66666-ABG66758 represent human
CC novel polypeptides of the invention.
XX Sequence 1296 AA;

Query Match 9.2%; Score 597; DB 23; Length 1296;
Best Local Similarity 24.1%; Pred. No. 1.2e-22;
Matches 344; Conservative 203; Mismatches 464; Indels 414; Gaps 67;

QY 12 LSIASAKTTTGVDSF-----NSLTWSNAANYAFKPGPYTNVAVLQWSLDGT-----SA 61
Db 21 VSMTSA-TIPSVRPTFTSHNTLITSLTLTF---PGTYSFSSMSASDGTHTHTTITS 75
QY 62 NPGDFTLNMPC-----VFKYTTTSOT-----SVDLTADGVKATATCQYSGEEFTTF 107
Db 76 LPASTSLHTAESTTAHTTTTSTFTTWTNESPSSSVATSTG---QTFSSSTATFTET 132
QY 108 STLTCTVNDALKSSIKAFCTVTLPIAFNVYGGSTDLSDSKCFTAGTNTVTFNDGDKOI 167
Db 133 TLTPT-TDFSEETLTAMTSTPTTSSITPTNTVTSMTMTSWPATNTLS-----183
QY 168 STDVFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGMTGSSSSGNDVAIDC 227
Db 184 SLTTNLTSPVPSTERTTSH-TNINPVSTL-----V 215
QY 228 SNHIGITKG--LNDWNPVSSSESFYTKTSNGIQIKYQNVNYPAGYRPFIDAYISATDV 285
Db 216 TLTPTTITRSTPTSETTYPISS-----TSTVSTTEITY-----250
QY 286 NOYTLAYTNDYTCAGSRLOSKPFTLRWTGYKNKSDAGSNGIVIVATRTVTDSTAVTTLP 345
Db 251 -----STTWTETSSSATSLP 265
QY 346 F-NPSVDKTKTIE----ILOPIPTTITTSYGVVTSYLTATKAPIGETA-----TVIVDVP 396
Db 266 LTPSLVSTTETAKTPTTL--VTTTTKTSH--STTSST--VYSTASTHTTATSV 319
QY 397 YHTTTTSTSEWGTITTTTTRTNPTDSIDTV-----VQVPLPNPNTVSTYKWSQSFAIT 451
Db 320 TTLGTMTVS--TSRIPSTVSTSIPTSQPKTVNSSSGGITGSLPMMDLTLSGY-----TV 371

Qy	1082	-----PSMGENGLTSTIEIATNTSP-----TEAPSPAVSS-----GTDVTEPT	1122
Db	991	FSVISLSPCTCITITIVPASPTDPCVEMDPSTEATSPPTTPLTVFPFTEMTCTPT	1050
Qy	1123	DTRQPTLTSTKTNSELVATQATNENGGKSPSTDL--TSSLTGTCTASTS-----ANS	1176
Db	1051	SISIQ-TLLTYMTSSMPPESSISPNASSSTGCTGVTNTVFTSTRLPTSETWLSNS	1109
Qy	1177	ELV-----TSGSVTGAVASANDQSHSTSVTNSNSIVSNTPTTILSQ	1220
Db	1110	SVIPLPLPGVSTIPLTMKPSSSPLTILRTSSKSTHSPPTTRTSEIPVATQPTILTSR	1169
Qy	1221	VTS-----SSPSTNTFIATYDGSIIQHSWLYGLITLLSLF	1259
Db	1170	RTTRITSQMTTQSLTTTAGTCDNNG-----TWEQGCACLPGF	1208
RESULT 9			
AA	AY95559	standard; Protein; 2870 AA.	
XX	AC	AY95559;	
XX	DT	10-OCT-2000 (first entry)	
XX	DE	Caenorhabditis elegans LOV-1 deletion allele lov-1(sy582del).	
XX	KW	Location of vulva; LOV-1 gene; nematode; mating behaviour;	
KW	KW	polycystin; polycystic kidney disease; animal model;	
KW	KW	signal transduction; mutant; mutein.	
XX	OS	Caenorhabditis elegans.	
OS	OS	Synthetic.	
XX	PN	WO2000040711-A2.	
XX	PD	13-JUL-2000.	
XX	PF	06-JAN-2000; 2000WO-US00521.	
XX	PR	06-JAN-1999; 99US-0115127.	
XX	PA	(CALY) CALIFORNIA INST OF TECHNOLOGY.	
XX	PI	Sternberg PW, Barr MM;	
XX	DR	WPI; 2000-452537/39.	
XX	PT	Isolated nucleic acid molecules from Caenorhabditis elegans useful for	
PT	PT	producing transgenic nematodes with altered mating behavior for	
PT	PT	identifying genes or regulatory factors involved in polycystic kidney	
PT	PT	disease -	
XX	PS	Example 1; Page 131-139; 142pp; English.	
XX	CC	The present sequence is that of a deletion allele, termed	
CC	CC	lov-1(sy582del), of the Caenorhabditis elegans LOV-1 protein	
CC	CC	(see AAY95556), generated by genomic deletion of the PKD/channel	
CC	CC	domain of the lov-1 gene by PCR. LOV-1 is an orthologue of	
CC	CC	the human polycystin-1 (PKD1) gene that is defective in human	
CC	CC	autosomal dominant polycystic kidney disease (PKD). The LOV-1	
CC	CC	protein is involved in chemosensory or mechanosensory signal	
CC	CC	transduction in sensory neurons. It is required for 2 male	
CC	CC	sensory behaviours, 'response' of males to hermaphrodites, and	
CC	CC	'location' of the vulva. Nematodes such as C. elegans that	
CC	CC	express mutant or wild-type LOV-1 or PKD-2 (see AAY95557) can be used	
CC	CC	to study the functions of the proteins encoded by these genes, to	
CC	CC	screen for other genes involved in PKD, to identify mutations	
CC	CC	involved in the disease, and to screen for drugs that affect PKD.	
CC	CC	Behaviours controlled by the action of the genes or gene products	
CC	CC	are identified and used in the assays. Hence, an animal model is	
CC	CC	provided that permits study of the etiology of PKD and provides a	
CC	CC	tool to identify the genes involved in the disease pathway.	

Sequence 957 AA;

Query Match		8.4%; Score 545.5; DB 21; Length 957;
Best Local Similarity		25.9%; Pred. No. 3.8e-20;
Matches 266; Conservative 139; Mismatches 406; Indels 215; Gaps 43;		
Qy	338	TTAVTTLPFFNSVDKTKTIELQPIPTTTITTSVGVTTSTLYLTKTAPIGETATVIVDVPY 397
Db	6	TTAPFGSTMPGVQSESTAS--HSSPGSTDTLTSPGSTTA-----SSLGPESTTFHSPG 58
Qy	398	HTTTTIVSEMT---GTITTTTRTNPTDSIDTVVQVPLPNPTVSTT-----EWSQ 446
Db	59	STETTLDPDNTASGLLEASTPVHSGSPHTL-----SPAGSTFQGESSTTFQSPN 112
Qy	447	SFATTTVTAPPGTDTVIRPPNH---TVTTEYWSQSFAT-----TTVTAPPGGT 497
Db	113	SKDTT---PAPPTTSAFVELSTSHGSPSPSTPTTFHSASSTTLGRSEESTVHSSPVAT 169
Qy	498	DSVIIRPPNPTVTTTEWSQSFATTTVTAPPGTDSVIRPPNPTVTTTEWSQSYA 557
Db	170	ATP-----PSPARSTT---SGLVEESTYHSGPGSTQTMHPPESTTSGRGEESTTSHS 221
Qy	558	TTTTVTAPPGTDSVIRPP---PNHTVTT-----TEYWSQSYA---TTTT 597
Db	222	TTHTISSAP-STTSALVEEPTSYHSSPGSTATTFPDSSTTSGRSEESTASHNQDATGT 280
Qy	598	VTAPPGTDTVIRPPNHVTTTEYWSQSFATTT---TTVTGPPSGTDTVIRPPNPT 653
Db	281	IVLPARSTSVLLGES-----TSPISGSMETTALPGSTTTPGLSEKSTTFHSSPRSPA 335
Qy	654	VTTTEYWSQYAT---TTTITAPGCTDTVIRPPNHVTTTEYWSQSYATTTVTAPP 710
Db	336	TTLSPASTSGVSEESTSHSRPGSTHTTAPDPS-----TTTPGLSRH---STTSHSP 387
Qy	711	GETDTVILIREPPNHTVTTTEYWSQYATTTVTAPPGCTDTVIRPPNPTVTTTEYWSQ 770
Db	388	GSTDITLL--PAS---TTTSPQSE---STTSHSGPGSTDAL-----SPGSTALSFGQ 434
Qy	771	SFATTTVTAPPGTDTVIRPPNHVTTTEYWSQSFATTT---TTVTGPPSGTDTVIRPPNPT 822
Db	435	E--STTFHSPGSTHTTLPDSTTSSGIVAEVSHSGSTGSPRTTLPASSTSPGLQGE 491
Qy	823	TSDLSTPSSSMNPTTSSDCMLLSSTLTVESETTTELLICSDGKCSRLSSSGIVTN 882
Db	492	STAFQTHPASTHTTPTST-----PSTATAPVEESTYHRSPPSTPTTFPEASP---TT 540
Qy	883	PDSNESSIVTGVPTAS--TWSDLSLSTGDIGASATSSDNVSKSGVSTTETSTVITQTPN 940
Db	541	SGHSEKSTIFHSSPDAGSTTTPSSAHSTSGRGESTTSRISPGSTETTLPGST---TPG 597
Qy	941	PLSSSVTSLAQSLIPSVSESESKVFTTSGNDQSGHDQSTSEIBIVTSSTKVLPP 1000
Db	598	-LSEASTTFYSSPRSPPTTLSPASMTSLGVE-ESTTSRQSGSTHVSVPASTI---TP 652
Qy	1001	VVS--SNVTLTSETNTR-----EQTTL-----STTSNITIEDITS--- 1036
Db	653	GLSEESTTVYSSPGSTETTFVPRSTTTSVARGEPTTFHSPASTHTTLFTEDSTTSLGT 712
Qy	1037	-OPTGDNQDNSTSNVPTVATSLTASASEDNK-SGSHESASTSLKPMGENSEGL---- 1090
Db	713	EESTAFPGSPASTQGLP--AALTADLGEESTTFPSSSGSTGTTLSPARTTSLGVEES 770
Qy	1091	-----TTSTEIEAFTT-----SPTAPSPAVSSGTDVTTETPTDTR 1125
Db	771	TPSRLSPSTETTLPGSPPTTSLSEKSTTFTVSPRSPDALTSPATTTSSGVSEESSH 830
Qy	1126	EQPTTLSTTS-----KTSELVATQATNENKGPSTDLTSSL----- 1164
Db	831	SQPGSTHTTAPDSTTTTSLGQEPKTHSSQSGTEAT-----LSPGSTTASLGGQOSTF 885
Qy	1165	--TTGTSASTSANELVTSVSGVAVASANDQSHSTSVTNSNIVSNTPQTLSQQVT 1222
Db	886	HSSPGDITETLLPDDTTFYSGLVEASTPHTSHSTGSLHTLTLPASSTISAGLQESTTFQSWP 945

Qy 1223 SSSPST 1228
Db 946 SSSDTT 951

RESULT 12
ID AAM24513 standard; Protein; 957 AA.

AC AAM24513;

DT 12-OCT-2001 (first entry)

XX C900P predicted amino acid sequence.

DE Human; immunotherapy; diagnosis; colon cancer; colon tumour;

XX immunogenic; gene therapy; vaccine; colonic cancer.

OS Homo sapiens.

PN WO200149716-A2.

XX 12-JUL-2001.

PF 29-DEC-2000; 2000WO-US35596.

XX 30-DEC-1999; 99US-0476296.

PR 10-JAN-2000; 2000US-0480321.

PR 15-FEB-2000; 2000US-0504629.

PR 06-MAR-2000; 2000US-0519444.

PR 19-MAY-2000; 2000US-0575251.

PR 29-JUN-2000; 2000US-0609448.

PR 28-AUG-2000; 2000US-0649811.

XX (CORI-) CORIAX CORP.

XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;

PI King GE, Wang T, Jiang Y;

XX WPI; 2001-441847/47.

XX Colon tumor associated proteins and nucleic acids useful for the

PT prevention, diagnosis and treatment of colonic cancer -

XX Claim 2; Page 437-440; 472pp; English.

PS The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantify the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.

XX Sequence 957 AA;

Search completed: June 11, 2003, 17:13:28
Job time : 58 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:11:31 ; Search time 19 Seconds
(without alignments)
2750.535 Million cell updates/sec

Title: US-09-715-876-8

Perfect score: 6495

Sequence: 1 MLQQTLLFLYLSIASAKTI.....SIIOHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match %	ID	Description
1	6465	99.5	1260	1 ALS1_CANAL
2	4204.5	64.7	1119	1 ALS3_CANAL
3	3194	49.2	1419	1 ALA1_CANAL
4	1992.5	30.7	468	1 ALS2_CANAL
5	1819	28.0	469	1 ALS4_CANAL
6	940	14.5	1537	1 FLO1_YEAST
7	932.5	14.4	1322	1 YAG3_YEAST
8	721.5	11.1	5179	1 MUC2_HUMAN
9	711	10.9	1075	1 FLO5_YEAST
10	675.5	10.4	1367	1 AMYH_YEAST
11	577	8.9	1161	1 DAN4_YEAST
12	553.5	8.5	3178	1 YS89_CAEEL
13	547.5	8.4	1609	1 FIG2_YEAST
14	521.5	8.0	1140	1 YN96_YEAST
15	501	7.7	1169	1 YK82_YEAST
16	464.5	7.2	670	1 VG50_HSV11
17	446	6.9	1802	1 HKR1_YEAST
18	443.5	6.8	1251	1 YQJ3_CAEEL
19	418.5	6.4	725	1 AGAL_YEAST
20	415.5	6.4	1306	1 MSB2_YEAST
21	409.5	6.3	1041	1 EGT2_YEAST
22	406	6.3	2090	1 HFC1_MESAU
23	404.5	6.2	2035	1 HFC1_HUMAN
24	392	6.0	797	1 VGLX_HSVB
25	374.5	5.8	881	1 YJH8_YEAST
26	353	5.4	650	1 SAG1_YEAST
27	351.5	5.4	5376	1 ZAN_MOUSE
28	349.5	5.4	860	1 CH12_COCIM
29	347.5	5.4	1150	1 APWU_PIG
30	338	5.2	1848	1 CBPA_CLOCL
31	337	5.0	2358	1 YEEJ_ECOLI
32	323	5.0	907	1 VGP3_EBV
33	320	4.9	886	1 VGP3_EBVA8

ALIGNMENTS

RESULT 1
ALS1_CANAL
ID ALS1_CANAL STANDARD; PRT; 1260 AA.
AC P46590;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Agglutinin-like protein 1 precursor.
GN ALS1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11651 / B792;
RA Hoyer L.L., Scherer S., Shatzman A.R., Livi G.P.;
RT "Candida albicans ALS1: domains related to a Saccharomyces cerevisiae sexual agglutinin separated by a repeating motif.";
RL Mol. Microbiol. 15:39-54(1995).
CC 1- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC 1- SIMILARITY: TO YEAST SAG1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L25902; AAC1649.2;
KW Cell adhesion; Glycoprotein; Repeat; Signal.
FT SIGNAL 1 17
FT CHAIN 18 1260 AGGLUTININ-LIKE PROTEIN 1.
FT DOMAIN 433 792 10 X 36 AA TANDEM REPEATS.
FT REPEAT 433 468 1-1.
FT REPEAT 469 504 1-2.
FT REPEAT 505 540 1-3.
FT REPEAT 541 576 1-4.
FT REPEAT 577 612 1-5.
FT REPEAT 613 648 1-6.
FT REPEAT 649 684 1-7.
FT REPEAT 685 720 1-8.
FT REPEAT 721 756 1-9.
FT REPEAT 757 792 1-10.
FT DOMAIN 983 1152 2 X 26 AA APPROXIMATE REPEATS.
FT REPEAT 983 1043 2-1.
FT REPEAT 1092 1152 2-2.
FT DOMAIN 399 404 POLY-THR.
FT DOMAIN 408 418 POLY-THR.
FT DOMAIN 450 455 POLY-THR.
FT DOMAIN 486 491 POLY-THR.
FT DOMAIN 522 527 POLY-THR.

34 312 4.8 2660 1 YEEJ_ECO57 Q8x8v7 escherichia
35 311 4.8 1253 1 DSEP_HUMAN Q9nzw4 homo sapien
36 305 4.7 2090 1 N214_HUMAN P35658 homo sapien
37 304.5 4.7 3396 1 PGCY_HUMAN P13611 homo sapien
38 302 4.6 395 1 YIQ9_YEAST P40442 saccharomyc
39 302 4.6 3562 1 PGCY_CHICK Q90953 gallus gall
40 299.5 4.6 3381 1 PGCY_BOVIN P81282 bos taurus
41 295.5 4.5 662 1 MUC1_XENLA Q05049 xenopus lae
42 294 4.5 1849 1 IGA4_HAEIN P45386 haemophilus
43 293.5 4.5 610 1 MUC4_HUMAN Q99102 homo sapien
44 293 4.5 1399 1 N121_RAT P52591 rattus norv
45 286 4.4 636 1 YNR6_YEAST P53882 saccharomyc

FT	DOMAIN	558	563	POLY-THR.	
FT	DOMAIN	594	599	POLY-THR.	
FT	DOMAIN	630	635	POLY-THR.	
FT	DOMAIN	666	671	POLY-THR.	
FT	DOMAIN	702	707	POLY-THR.	
FT	DOMAIN	738	743	POLY-THR.	
FT	DOMAIN	774	779	POLY-THR.	
FT	DOMAIN	874	877	POLY-SER.	
FT	CARBOHYD	471	471	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	579	579	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	615	615	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	687	687	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	723	723	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	820	820	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	886	886	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	918	918	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	973	973	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1045	1045	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1068	1068	N-LINKED (GLCNAC. .)	(POTENTIAL).
SQ	SEQUENCE	1260 AA;	132641 MW;	763D1063A2354C24	CRC64;
Query Match 99.5%; Score 6465; DB 1; Length 1260;					
Best Local Similarity 99.8%; Pred. No. 1.6e-270;					
Matches 1255; Conservative 0; Mismatches 5; Indels 0; Gaps 0;					
QY	1	MLQOFTLLFLYLISIAKAITGVDFDSFNSLTWSNAANYAFKPGYPTWNAVLGWSLDGTS	60		
DB	1	MLQOFTLLFLYLISIAKAITGVDFDSFNSLTWSNAANYAFKPGYPTWNAVLGWSLDGTS	60		
QY	61	ANPGDFTFLNMPGVKYYTSQTSVDLTADGVKYATCQFYSGEFTTFTLTCTVNDALKS	120		
DB	61	ANPGDFTFLNMPGVKYYTSQTSVDLTADGVKYATCQFYSGEFTTFTLTCTVNDALKS	120		
QY	121	SIKAFGTVLPATFNNGGTSSTDLSDSKCFAGTNTVTFNDGDKDISIDVFEKSTVDP	180		
DB	121	SIKAFGTVLPATFNNGGTSSTDLSDSKCFAGTNTVTFNDGDKDISIDVFEKSTVDP	180		
QY	181	SAYLYASRVMSLKNVTTFLFVAPQCENGYTSGMTGFSNGDVAIDCSNIHIGITKGLND	240		
DB	181	SAYLYASRVMSLKNVTTFLFVAPQCENGYTSGMTGFSNGDVAIDCSNIHIGITKGLND	240		
QY	241	WNPVSESEFSYKTKTSNGIQIKYONVAGYRPFIDAYISATDVNQYTLAYTNDYTCAG	300		
DB	241	WNPVSESEFSYKTKTSNGIQIKYONVAGYRPFIDAYISATDVNQYTLAYTNDYTCAG	300		
QY	301	SRLOKPFLLRWGTYGKNSDAGNGIVIVATRTVTDSTTAVTLPENPSVDKTKTEILQ	360		
DB	301	SRLOKPFLLRWGTYGKNSDAGNGIVIVATRTVTDSTTAVTLPENPSVDKTKTEILQ	360		
QY	361	PIPTTTTTSYGVVTSYLTKTAPIGETATVIVDVPYHTTIVTSEWGTITTTTTRTP	420		
DB	361	PIPTTTTTSYGVVTSYLTKTAPIGETATVIVDVPYHTTIVTSEWGTITTTTTRTP	420		
QY	421	TDSIDVNVQVPLPNFTVSTTEYWSQSFATTTTVPAPGDTDVIIREPNNHTVTTTEY	480		
DB	421	TDSIDVNVQVPLPNFTVSTTEYWSQSFATTTTVPAPGDTDVIIREPNNHTVTTTEY	480		
QY	481	SQSFAITTTTVPAPGDTDVIIREPNNHTVTTTEYWSQSFATTTTVPAPGDTDVII	540		
DB	481	SQSFAITTTTVPAPGDTDVIIREPNNHTVTTTEYWSQSFATTTTVPAPGDTDVII	540		
QY	541	PPNPTVTTEYWSQSFATTTTVPAPGDTDVIIREPNNHTVTTTEYWSQSFATTTT	600		
DB	541	PPNPTVTTEYWSQSFATTTTVPAPGDTDVIIREPNNHTVTTTEYWSQSFATTTT	600		
QY	601	PPGCTDVIIREPNNHTVTTTEYWSQSFATTTTVPAPGDTDVIIREPNNHTVTT	660		
DB	601	PPGCTDVIIREPNNHTVTTTEYWSQSFATTTTVPAPGDTDVIIREPNNHTVTT	660		
QY	661	SQSFAITTTTVPAPGDTDVIIREPNNHTVTTTEYWSQSFATTTTVPAPGDTDVII	720		
DB	661	SQSFAITTTTVPAPGDTDVIIREPNNHTVTTTEYWSQSFATTTTVPAPGDTDVII	720		

Qy	721	PPNHTVTTEYWSQSYATTTTVPAPGDTDVIIREPNNPTVTTEYWSQSFATTTTVA	780
Db	721	PPNHTVTTEYWSQSYATTTTVPAPGDTDVIIREPNNPTVTTEYWSQSFATTTTVA	780
Qy	781	PPGCTDVIIYESMSSSKISTSNIDITSIIIPSPRPHYVNSTSDISTFESSMNTPTSI	840
Db	781	PPGCTDVIIYESMSSSKISTSNIDITSIIIPSPRPHYVNSTSDISTFESSMNTPTSI	840
Qy	841	SSDGMLLSSTTLVTESETTTTELICSDGKCSRSLSSSGIVTNPDSNESSIVTSTVPTAST	900
Db	841	SSDGMLLSSTTLVTESETTTTELICSDGKCSRSLSSSGIVTNPDSNESSIVTSTVPTAST	900
Qy	901	MSDSLSTDGISATSSDNVSKGVSTTETSVTTIQTTPNPLSSSVTSLTQLSSIPVSE	960
Db	901	MSDSLSTDGISATSSDNVSKGVSTTETSVTTIQTTPNPLSSSVTSLTQLSSIPVSE	960
Qy	961	SESKVFTTNGDNQSGHDSQSSTETIEIVTTSSTKVLPPVSSNTDLTSEPTNTRQPT	1020
Db	961	SESKVFTTNGDNQSGHDSQSSTETIEIVTTSSTKVLPPVSSNTDLTSEPTNTRQPT	1020
Qy	1021	TLSTTNSITEDITTSQPTGDNCDNTSSTNPVPTVATSTLASAEEENKSGSHESASTSL	1080
Db	1021	TLSTTNSITEDITTSQPTGDNCDNTSSTNPVPTVATSTLASAEEENKSGSHESASTSL	1080
Qy	1081	KPSMGNSGLTSTTEIATTTSTPEAPSPAVSGTDVTTPTDTRQPTTLSTTSKTNSE	1140
Db	1081	KPSMGNSGLTSTTEIATTTSTPEAPSPAVSGTDVTTPTDTRQPTTLSTTSKTNSE	1140
Qy	1141	LVATTQATNENGKSPSTDLTSSLTGTSTASTSANSELVTSVSGVTGGAVASANDQSHST	1200
Db	1141	SVATTQATNENGKSPSTDLTSSLTGTSTASTSANSELVTSVSGVTGGAVASANDQSHST	1200
Qy	1201	SVTNSNSIVNTPQTTLSSQOVTSSSPSTNTFTASTYDGSIIQHSWTWLYGLITLISLFI	1260
Db	1201	SVTNSNSIVNTPQTTLSSQOVTSSSPSTNTFTASTYDGSIIQHSWTWLYGLITLISLFI	1260
RESULT 2			
ALS3_CANAL STANDARD; PRT; 1119 AA.			
ID	ALS3_CANAL		
AC	074623;		
DT	15-JUL-1999 (Rel. 38, Created)		
DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DT	15-JUL-1999 (Rel. 38, Last annotation update)		
DE	Agglutinin-like protein 3 precursor.		
GN	ALS3.		
OS	Candida albicans (Yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.		
OX	NCBI_TaxID=5476;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=1161;		
RX	MEDLINE=98309840; PubMed=9644209;		
RA	Hoyer L.L., Payne T.L., Bell M., Myers A.M., Scherer S.;		
RT	"Candida albicans ALS3 and insights into the nature of the ALS gene family.";		
RL	Curr. Genet. 33:451-459(1998).		
CC	-!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.		
CC	-!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; U87956; AAC39486.1;		
KW	Cell adhesion; Glycoprotein; Repeat; Signal.		
FT	SIGNAL	1	117
			POTENTIAL.

```
FT DOMAIN 433 792 10 X 36 AA TANDEM REPEATS.
FT REPEAT 433 468 1-1.
FT REPEAT 439 504 1-2.
FT REPEAT 505 540 1-3.
FT REPEAT 541 576 1-4.
FT REPEAT 577 612 1-5.
FT REPEAT 613 648 1-6.
FT REPEAT 649 684 1-7.
FT REPEAT 685 720 1-8.
FT REPEAT 721 756 1-9.
FT REPEAT 757 792 1-10.
FT DOMAIN 399 404 POLY-THR.
FT DOMAIN 450 455 POLY-THR.
FT DOMAIN 557 563 POLY-THR.
FT DOMAIN 593 597 POLY-THR.
FT DOMAIN 630 635 POLY-THR.
FT DOMAIN 666 671 POLY-THR.
FT DOMAIN 702 707 POLY-THR.
FT DOMAIN 738 743 POLY-THR.
FT DOMAIN 774 777 POLY-THR.
FT DOMAIN 1044 1047 POLY-THR.
FT CARBOHYD 471 471 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 543 543 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 579 579 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 651 651 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 687 687 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 723 723 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 845 845 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 987 987 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1119 AA; 119927 MW; 6A3FB3FC8C879A71 CRC64;

Query Match 64.7%; Score 4204.5; DB 1; Length 1119;
Best Local Similarity 64.9%; Pred. No. 1.2e-173;
Matches 844; Conservative 106; Mismatches 128; Indels 223; Gaps 18;

QY 1 MLOQFTLLFLYLISAKTITGVDFNSLWSNAANYAFKPGGYPTNNAVIGWSLDGTS 60
DB 1 MLOQFTLLIYLSVATATITGVFNSLWSNAATYKYGPGPTNNAVIGWSLDGTS 60

QY 61 ANPGDTFTLNMPCVFKYTTTSQTSVDLTADGVKATCQFYSGEFTFTSLTCTVNDALKS 120
DB 61 ASPGDTFTLNMPCVFKYTTTSQTSVDLTAGVRYATCQFOAGEFTFTSLTCTVSNLTP 120

QY 121 SIKAGTFTLPAFNVGGTSGTSDLEDKSCFTAGTNTVTFNDGDKDISIDVFEKSTVDP 180
DB 121 SIKALGTFTLPAFNVGGTSGTSDLEDKSCFTAGTNTVTFNDGKKISINVDERSNDP 180

QY 181 SAYLASRVPMLNKVTLTFVAPQCENGYTSCTMGFSSNGDVAIDCSNIHIGITKGLND 240
DB 181 KGLYLSRVIPSLNKVTLTFVAPQANGYTSCTMGFANTYGVQIDCSNIHIGITKGLND 240

QY 241 WNPVSSSEFSYTKTCTSNQIOIKYQNPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
DB 241 WNPVSSSEFSYTKTCSNGIFITYKNVPAGYRPFVDAVISATDVNSYTLSEANEYTCAG 300

QY 301 SRLQSKPTLRTWGYKNSDAGSNGIVVATRTVTDSTAVTTLTLPNPNVDKTKTIELQ 360
DB 301 GYWRAPFTLRTWGYRNSDAGSNGIVVATRTVTDSTAVTTLTLPNDRKTKTIELK 360

QY 361 PIPTTTITTSYGVGTVTLTKAPIGETATVVDVPYHTTNTVTSEWGTCTITTTTRNP 420
DB 361 PIPTTTITTSYGVGTVSTKAPIGETATVVDVPYHTTNTVTSEWGTCTITTTTRNP 420

QY 421 TDSIDTVVQVPLNPFTVSTTEYWSQSFATTTTAPPGETDVLIIREPNNHTVTTTVA 480
DB 421 TDSIDTVIVQVSPNPFTVSTTEYWSQSFATTTTITPGCNTDVLIIREPNNHTVTTTVA 480

QY 481 SOSFATTTTVAAPPGETDVLIIREPNNHTVTTTTEYWSQSFATTTTAPPGETDVLII 540
DB 481 SESYTTTSTTAPPGETDVLIIKEPNNHTVTTTTEYWSSEYTTTSTTAPPGETDVLII 540
```

```
QY 541 PPNPTVTTTTEYWSQSFATTTTAPPGETDVLIIREPNNHTVTTTTEYWSQSFATTTTVA 600
DB 541 PPNHTVTTTTEYWSQSYTTTAPPGETDVLIIREPNNHTVTTTTEYWSQSYTTTVA 600
QY 601 PPGGTDVLIIREPNNHTVTTTTEYWSQSFATTTTAPPGETDVLIIREPNNHTVTTTTEY 660
DB 601 PPGGTDVLIIREPNNHTVTTTTEYWSQSYTTTAPPGETDVLIIREPNNHTVTTTTEY 660
QY 661 SOSYATTTTITAPPGETDVL-----IRE 684
DB 661 SOSYATTTTITAPPGETDVLIIREPNNHTVTTTTEYWSQSFATTTTAPPGETDVLII 720
QY 685 PPNHTVTTTTEYWSQSFATTTTAPPGETDVLIIREPNNHTVTTTTEYWSQSFATTTTVA 744
DB 721 PPNHTVTTTTEYWSQSFATTTTAPPGETDVLIIREPNNHTVTTTTEYWSQSFATTTTVA 780
QY 745 PPGGTDVLIIREPNNHTVTTTTEYWSQSFATTTTAPPGETDVLIIREPNNHTVTTTTEY 804
DB 781 PPGGTDVLIIREPNNHTVTTTTEYWSQSYTTTAPPGETDVLIIREPNNHTVTTTTEY 836
QY 805 DITSIIPSRPHYVNSITSDLSSTFESSMNTPTSISSDGMLLSSTLTTESETTELIC 864
DB 837 -----SPSRPHYTNHTT-----LWSTTWIETKITITETSC 866
QY 865 SDGKCCSRLLSSSGIVTPDSSNESSIVTPTASTMSDLSSTDGI SATSSDNVSKGV 924
DB 867 EGDGCGSVSVSTRIVTPNNIETPMVTNTVSTTTESTS-QSPSCI-----FSESGV 918
QY 925 SVTTETS-VTTIQTTPNPLSSSVTSLTQLSSPSVSESESKVTFTSNGDNQSGTHDSQST 983
DB 919 SVTESSTVTTAQTN-----PSVPTTESEVVFTTKGNENGPEYSPST 961
QY 984 STEIELVTTSSKVLPPVSSNTDLTSEPTNTREQPTTLTSTNSITEDITTSQPTGDCNG 1043
DB 962 N-----VKSMSENSEFT----- 974
QY 1044 DNTSSTNPVPTVATSTLASAEEDNKGSHESASTSLKPSMGENSELTTSTEIE---ATT 1100
DB 975 -----TSTAAS-----TSTDIENETIAT 992
QY 1101 TSPTAPSPAVSGTDVTTTEPTDTRQPTTLTSTKTNSELVATQATNEN-GGKSPSTD 1159
DB 993 TGSVEASSPIISSADET-----TTVTTTAEASTSVI---EOPTNNGGKAPSA- 1038
QY 1160 LRTSLTTGTSAITSANSELVTSVYTGGAASANDQSHSTSVTNSINSVNTPTQLSQ 1219
DB 1039 -----TSSPSTTTTANNDSVITG-----TTSTNQSQSOYIN-----SDTQPTLSQ 1080
QY 1220 QVTSSPSTNTFIATYDGSSTIOHSTWLYGLITLLSLFI 1260
DB 1081 QMTSSILVSLH--MLTTFDGGSGVIOHSTWLCGLITLLSLFI 1119
```

```
RESULT 3
ALAL_CANAL
ID ALAL_CANAL STANDARD; PRT; 1419 AA.
AC O13368;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Agglutinin-like protein ALAL precursor (Agglutinin-like adhesin).
GN ALAL OR ALS5.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
RN NCBI_TaxID=5476;
RP SEQUENCE FROM N.A.
RX MEDLINE=98053977; PubMed=9393828;
RA Gaur N.K., Klotz S.A.;
RT "Expression, cloning, and characterization of a Candida albicans
RT gene, ALAL1, that confers adherence properties upon Saccharomyces
```

RT cerevisiae for extracellular matrix proteins.*;
 RL Infect. Immun. 65:5289-5294(1997).
 CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
 CC -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF025429; AAB88883.1;
 KW Cell adhesion; Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 1419 AGGLUTININ-LIKE PROTEIN ALA1.
 FT DOMAIN 399 404 POLY-THR.
 FT DOMAIN 408 418 POLY-THR.
 FT DOMAIN 437 441 POLY-THR.
 FT DOMAIN 673 676 POLY-SER.
 FT DOMAIN 687 690 POLY-SER.
 FT DOMAIN 700 703 POLY-SER.
 FT DOMAIN 719 724 POLY-SER.
 FT DOMAIN 749 752 POLY-SER.
 FT DOMAIN 787 791 POLY-SER.
 FT DOMAIN 869 872 POLY-SER.
 FT DOMAIN 875 883 POLY-SER.
 FT DOMAIN 901 911 POLY-SER.
 FT DOMAIN 1216 1221 POLY-SER.
 FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1301 1301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1326 1326 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1419 AA; 149635 MW; 249F33F688A9D5B6 CRC64;

Query Match 49.2%; Score 3194; DB 1; Length 1419;

Best Local Similarity 49.5%; Pred. No. 3.6e-130;

Matches 708; Conservative 182; Mismatches 355; Indels 186; Gaps 29;

QY 1 MLOQFTLLFLYLISAKITGVDFNSLTWNAANYAFKPGYPTWNAVLGWSLDGYS 60
 DB 1 MLOQFTLLFLYLISAKITGVDFNSLTWNAANYAFKPGYPTWNAVLGWSLDGYS 60
 QY 61 ANPGDFTLLMPCVFKYKTSQTSVDLTADGVKYATCFYSGEEFTTFLCTVNDALKS 120
 DB 61 ANPGDFTLLMPCVFKYKTSQTSVDLTADGVKYATCFYSGEEFTTFLCTVNNLRS 120
 QY 121 SIKAFGTVLPIAFNVGGTSGSDLEDSKCFYTAGTNTVTFNDGDKOISIDVEPEKSTVDP 180
 DB 121 SIKALGTVLPIAFNVGGTSGSDLEDSKCFYTAGTNTVTFNDGSKLSIAVNEPEKSTVDQ 180
 QY 181 SAYLYASRVMPSLNKVYTLFVAQPCNGYTSGMTGFSNGDVAIDCSNHHIGITKGLND 240
 DB 181 SGLYTSRFPMSLUNKATLVAQPCNGYTSGMTGFSNGDVAIDCSNHHIGISKGVND 240
 QY 241 WNPVSEFSYTKTCTSGNIOIKYQNPAGYRPFIDAXISADVAVQYTLAYNDYTCAG 300
 DB 241 WNPVTSSEFSYTKSCSFGISITYQNPAGYRPFIDAXISPSDNNOYQLSYKNDYTCVD 300
 QY 301 SRLQSPFTLRWTKYNSDAGSNGIVVATRTVSTTAVTLPNPSVDKTKTEILQ 360
 DB 301 DYWOHAPFTLRWTKYNSDAGSNGIVVATRTVSTTAVTLPNPSVDKTKTEILQ 360
 QY 361 PIPTTTITTSYGVVTSYLTAKTAPIGETATVIVDVPHYHTTTVTSEWGTITTTTRTP 420
 DB 361 PIPTTTITTSYGVVTSYLTAKTAPIGETATVIVDVPHYHTTTVTSEWGTITTTTRTP 420
 QY 421 TDSIDVWVQVPLNPVSTTEYSQSFAATTTVTAPPGGTDVRIIREPPNHTVTTEYW 480
 DB 421 TDSIDVWVQVPSNPNTTTTQWSESTSTTTITNSLKGTDSDSVIVREPPNHTVTTEW 480
 QY 481 SQSFATTTVTAPPGGTDSDSVIVREPPNHTVTTEYSQSFAATTTVTAPPGGTDSDSVI 540

DB 481 SESFATTTITSKPECTSDSVIVREPPNHTVTTEFWSEYATTTITNGPEGDSVIVRE 540
 QY 541 PNPPTVTTEYSQSFAATTTVTAPPGGTDSDSVIIREPPNHTVTTEYSQSFAATTTVTA 600
 DB 541 PNPPTVTTEFWSEYATTTITNGPEGTDSDSVIVKEPYNPTVTTEFWSEYATTTITN 600
 QY 601 PPGGTDVRIIREPPNHTVTTEYSQSFAATTTVTGPGGTDVRIIREPPNHTVT--- 657
 DB 601 GPEGTDSDSVIVREPPNHTVTTEFWSEYATTTITGPGGTDSDSVIHDPLESSSTAIE 660
 QY 658 -----EYWSQSFAATTTITAPPGGTDVLIIREPPNHTVT----- 691
 DB 661 SDSNINSSAQESSSVEQSSIVGLSSSDIPLSDMPSSSTGLTSSSTSVSSYDSD 720
 QY 692 -----TTEYSQS-----YATTTVTAPPGGTDVLIIREPPNHT 725
 DB 721 SSSIESSTLSSDRCSISIDTTFWSSSSDLESTITWSSSDAQSHLVQSSNSI 780
 QY 726 VTTTEYSQSFAATTTVTAPPGGTDVRIIREP-----PNPTVTTEYW----- 768
 DB 781 STSQELSSSSSESTF-----ATDALVSSDASSILSSDTSYSPSTISSDDPHTIA 835
 QY 769 -----SQSFATTT-----TVTAPPGGTDVRIIREPSSSKISTSSNDI----- 806
 DB 836 GESDLSISFITSTVEISSDSVSLTSDPASSFSSSSSLNSDSSSPSSDQSDILTSSES 895
 QY 807 TSIIPSEF-----RPHYVNSTSDLSFESSSMNTPTPISDGMLLSSTLTVTES 856
 DB 896 TLVVPFSLSSSSSLTYPHYVNSTYHASESSSVASPMASESAN---DDTYLTSES 953
 QY 857 ETTTELICDGKE---CSRLLSSSGIVTN-POSNESSIVTSTVPTASTMSDLSSTDG-I 911
 DB 954 TDTTSGIGTDSDSVITFCRDNGCCIVTGMPSIDSEGTSDVTTSFVASSTPFAEQ 1013
 QY 912 SATSSDNV---SKSGSVYTTTSTVTTIQTTPNPLSSSVTLQTLSSIPSSVSESKYFTFS 969
 DB 1014 SITDNPIDSSQTSASSSTKSSVSVDVTVNLSILLSETS---TLSSDDSTS-SDTSISST 1070
 QY 970 NGD-----NQGTHDSOSTSTETETIVTS---SPKVLPPVVSNTDLTSEPTNTR----- 1016
 DB 1071 NSDTGMINAGSHTSTASIKESSIQKTGVLSSYSLTSLSSDITIELITTELITTEL 1130
 QY 1017 -----EQPTTLSTNSITE-----DITTSQPTGNDNTSNTPVPTAT-----STLA 1061
 DB 1131 TTIEDNEPTFTSTPSSHSEIFSSDSDNSVLSKQVDRESTIKTSPTTDTVTVSSLSVHSEA 1190
 QY 1062 SA-----SEEDNKGSHESASTSLKPMGENSEGLTSTTELEATTTTPTAPSPAVSSGTD 1116
 DB 1191 STATLGENSEFNVASTPSNIATSLRSTSSSNHATSESGTVKSEASAEAIKPSPTS--TD 1248
 QY 1117 VTTEPTDTREQPTTLSTTKTN-----SELVATQATN---ENGKSPSTDLTSLTTGT 1168
 DB 1249 NRLSYSTEAKGITYANGSTNNLITESQVAAPTSDTSVLIENPVVTSITFDNSSAAVDQ 1308
 QY 1169 SASTSANSELY-----TSGSVTGAVASANDQSHSTSV-----TNSNSI 1208
 DB 1309 PSKTKIESIMNPDSTNETNNGFIATLSQAQVPSSEIHSSELISSTTTAKTTDASMGDSA 1368
 QY 1209 VSNTPOTTLISQVTSSTSPSTNTFIATDGSGLIOHSTWLYGLITLLSLF 1259
 DB 1369 ASNSQPTTLIQVATS--SYNOPLITYAGSSATKHPKSWLLKFIKFAVLF 1417
 RESULT 4
 ALS2_CANAL STANDARD; PRT; 468 AA.
 ID ALS2_CANAL
 AC 074657;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Agglutinin-like protein 2 precursor (fragment).
 GN ALS2.

OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RX MEDLINE=98440424; PubMed=9765564;
RA Hoyer L.L., Payne T.L., Hecht J.E.;
RT "Identification of Candida albicans ALS2 and ALS4 and localization of
RT als proteins to the fungal cell surface."
RL J. Bacteriol. 180:5334-5343(1998).
CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF024580; AAC64235.1; -;
KW Cell adhesion; Glycoprotein; Repeat; Signal.
FT SIGNAL 1 17
FT CHAIN 18 >468 AGGLUTININ-LIKE PROTEIN 2.
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 315 315 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 468 468
SQ SEQUENCE 468 AA; 50203 MW; BFE773E169ED0FAF CRC64;

Query Match 30.7%; Score 1992.5; DB 1; Length 468;
Best Local Similarity 30.2%; Pred. No. 3 5e-79;
Matches 376; Conservative 33; Mismatches 59; Indels 1; Gaps 1;

QY 1 MLQQTFLFLYLISASAKTITGVDFNSLTWSNAANYAFKPGYPTWNAVIGWSLDGTS 60
DB 1 MLQQLLLSLCVSVATAKVTGVFNSLTWANAASYPYRGPATPTAVIGWSLDGAT 60
QY 61 ANPGDTFTLNMPCEVKYTTTSQTSVDLTADGVKATCOFYSGEEFTFTSLCTVNDALKS 120
DB 61 ANPGDTFTLNMPCEVKFITDQTSVDLTAEGVKATCOFYSGEEFTFTSLKCTVSNLTLS 120
QY 121 SIKAGTFTLPIAFNVGGTSSDLEDSKCFAGTNTVTNFNDGDKDISIDVEFEKSTVDP 180
DB 121 SIKALGTFTLPIAFNVGGTSSVDLESSQCFAGTNTVTNFNDGDKKISIDVDFEKTNEA 180
QY 181 SAYLVASRVMPSLNKVTLFVAPQCENGYTSCTMGFSSNGDVAIDCSNIHIGITKGLND 240
DB 181 SGYFTASRLIPSINKVSIYVAPQACANGYTSAGMGFIVLTGDTTIDCSNVHVGITKGLND 240
QY 241 WNPVSSSEFSYTKTCTSNIGIQKONYPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
DB 241 WNPVSSDLSLNKTKCTSGTISITENVPAGYRPFIDVYTSVSGQNR-OLRYTNDYACVG 299
QY 301 SRLQSKPFTLRWTGKNSDAGSNGIVIVATTTVDSTAVTTLPLFPNSVDKTKTIEILQ 360
DB 300 SSLQSKPFLRLRGYNSEANSNGFIVATTTVDSTAVTTLPLFPNSVDKTKTIEILQ 359
QY 361 PIPTTTITTSYGVGVTSTLTAPIGETATVIVDVPYHTTTVTSEWGTGTTTTTRTNP 420
DB 360 PIPTTTITTSYGVGVTSTLTAPIGETATVIVDVPYHTTTVTSEWGTGTTTTTRTNP 419
QY 421 TDSIDTVVQVPLNPPTVSTTEYWSQSFATTTTVPAPGTDVVIIEP 469
DB 420 TDSIDTVVQVPSNPPTVTTEYWSQSYATTTTVPAPGTDVSIIEP 468

RESULT 5

ALS4_CANAL

ID ALS4_CANAL STANDARD; PRT; 469 AA.

AC 074660;

DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Agglutinin-like protein 4 precursor (Fragment).
GN ALS4.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RX MEDLINE=98440424; PubMed=9765564;
RA Hoyer L.L., Payne T.L., Hecht J.E.;
RT "Identification of Candida albicans ALS2 and ALS4 and localization of
RT als proteins to the fungal cell surface."
RL J. Bacteriol. 180:5334-5343(1998).
CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF024584; AAC64239.1; -;
KW Cell adhesion; Glycoprotein; Repeat; Signal.
FT SIGNAL 1 17
FT CHAIN 18 >469 AGGLUTININ-LIKE PROTEIN 4.
FT NON_TER 469 469
SQ SEQUENCE 469 AA; 49604 MW; OEDCAB19B89EFCB1 CRC64;

Query Match 28.0%; Score 1819; DB 1; Length 469;
Best Local Similarity 72.1%; Pred. No. 9 8e-72;
Matches 338; Conservative 48; Mismatches 83; Indels 0; Gaps 0;

QY 1 MLQQTFLFLYLISASAKTITGVDFNSLTWSNAANYAFKPGYPTWNAVIGWSLDGTS 60
DB 1 MLQQLLLSLCVSVATAKVTGVFNSLTWANAASYPYRGPATPTAVIGWSLDGAT 60
QY 61 ANPGDTFTLNMPCEVKYTTTSQTSVDLTADGVKATCOFYSGEEFTFTSLCTVNDALKS 120
DB 61 ASAGDTFTLDMPCVKFIFDQTSIDLAVDGRVATCNLSAEFTTFSVSVCTVTTMTFA 120
QY 121 SIKAGTFTLPIAFNVGGTSSDLEDSKCFAGTNTVTNFNDGDKDISIDVEFEKSTVDP 180
DB 121 DTKAIGTFTLPFSFSGSGSDVLANSCQCFAGTNTVTNFNDGDTISITVDPEKSTVAS 180
QY 181 SAYLVASRVMPSLNKVTLFVAPQCENGYTSCTMGFSSNGDVAIDCSNIHIGITKGLND 240
DB 181 SDRIILSRILPSLSQVNLFLPQECANGYTSCTMGFTAGTATDCSIVHVGISNGLND 240
QY 241 WNPVSSSEFSYTKTCTSNIGIQKONYPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
DB 241 WNPVSSSEFSYTKTCTSTSVLTQFONVPAGYRPFIDAYISATRVSSYTMQYNIYACVG 300
QY 301 SRLQSKPFTLRWTGKNSDAGSNGIVIVATTTVDSTAVTTLPLFPNSVDKTKTIEILQ 360
DB 301 AASVDDSFHTTWRGYSNSQAGSNGITIVVTTTVDSTAVTTLPLFPNSDTDKTKTIEILQ 360
QY 361 PIPTTTITTSYGVGVTSTLTAPIGETATVIVDVPYHTTTVTSEWGTGTTTTTRTNP 420
DB 361 PIPTTTITTSYGVGVTSTLTAPIGETATVIVDVPYHTTTVTSEWGTGTTTTTRTNP 420
QY 421 TDSIDTVVQVPLNPPTVSTTEYWSQSFATTTTVPAPGTDVVIIEP 469
DB 421 TDSIDTVVQVPSNPPTVTTEYWSQSYATTTTVPAPGTDVSIIEP 469

RESULT 6

FLO1_YEAST STANDARD; PRT; 1537 AA.
 AC P32768;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Flocculation protein FLO1 precursor (Flocculin 1).
 GN FLO1 OR YAR050W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=94262325; PubMed=8203162;
 RA Watari J., Takata Y., Ogawa M., Sahara H., Koshino S., Onnola M.-L.,
 RA Airaksinen U., Jaatinen R., Penttilae M., Keraenen S.;
 RT "Molecular cloning and analysis of the yeast flocculation gene FLO1.";
 RL Yeast 10:211-225(1994).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX MEDLINE=95249563; PubMed=7731988;
 RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
 RA Hall J., Ouellette B.F., Keng T., Barton A.B., Su Y., Davies C.K.,
 RA Storms R.K.;
 RT "The nucleotide sequence of chromosome I from Saccharomyces
 cerevisiae.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
 RN [3]
 RN PRELIMINARY SEQUENCE FROM N.A.
 RX MEDLINE=93289821; PubMed=8511970;
 RA Teunissen A.W.R.H., Holub E., van der Hucht J., van den Berg J.A.,
 RA Steensma H.Y.;
 RT "Sequence of the open reading frame of the FLO1 gene from
 Saccharomyces cerevisiae.";
 RL Yeast 9:423-427(1993).
 RN [4]
 RN REVIEW.
 RX MEDLINE=96076625; PubMed=7502576;
 RA Teunissen A.W., Steensma H.Y.;
 RT "Review: the dominant flocculation genes of Saccharomyces cerevisiae
 constitute a new subtelomeric gene family.";
 RL Yeast 11:1001-1013(1995).
 CC -1- FUNCTION: MAY BE DIRECTLY INVOLVED IN THE FLOCCULATION PROCESS.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 (potential).
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE FLOCCULIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X78160; CAA55024.1;
 DR EMBL: L28920; AAC09499.1; ALT_SEQ.
 DR PIR: S31230; S31230.
 DR SGD: S0000084; FLO1.
 DR InterPro: IPR001389; Flocculin.
 DR Pfam: PF00624; Flocculin; 18.
 KW Glycoprotein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 ? FLOCCULATION PROTEIN FLO1.
 FT PROPEP 274 1537 REMOVED IN MATURE FORM (POTENTIAL).
 FT DOMAIN 278 1087 SER/THR-RICH.
 FT DOMAIN 278 1087 18 X 45 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 278 322 1-1.
 FT REPEAT 323 367 1-2.
 FT REPEAT 368 412 1-3.

FT REPEAT 413 457 1-4.
 FT REPEAT 458 502 1-5.
 FT REPEAT 503 547 1-6.
 FT REPEAT 548 592 1-7.
 FT REPEAT 593 637 1-8.
 FT REPEAT 638 682 1-9.
 FT REPEAT 683 727 1-10.
 FT REPEAT 728 772 1-11.
 FT REPEAT 773 817 1-12.
 FT REPEAT 818 862 1-13.
 FT REPEAT 863 907 1-14.
 FT REPEAT 908 952 1-15.
 FT REPEAT 953 997 1-16.
 FT REPEAT 998 1042 1-17.
 FT REPEAT 1043 1087 1-18.
 FT DOMAIN 1118 1157 2 X 20 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 1118 1137 2-1.
 FT REPEAT 1138 1157 2-2.
 FT DOMAIN 1226 1392 3 X 51 AA APPROXIMATE REPEATS.
 FT REPEAT 1226 1266 3-1.
 FT REPEAT 1291 1341 3-2.
 FT REPEAT 1342 1392 3-3.
 FT DOMAIN 1408 1434 3 X 9 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 1408 1416 4-1.
 FT REPEAT 1417 1425 4-2.
 FT REPEAT 1426 1434 4-3.
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 262 262 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 374 374 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 419 419 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 554 554 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 599 599 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 644 644 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 689 689 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 734 734 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1537 AA; 160722 MW; 992650C6BE9A8CEA CRC64;
 Query Match 14.5%; Score 940; DB 1; Length 1537;
 Best Local Similarity 26.0%; Pred. No. 1.9e-33;
 Matches 417; Conservative 244; Mismatches 523; Indels 418; Gaps 85;
 QY 2 LQOFTLLFLVLSASAKT-----ITGVDFSE-----NSLTWSNAAYAF----- 40
 DB 10 LAVFTLLAL--TSVSGATEACLPAGORKSNNINFYQYSLKDSSTYSNAAYMAYGASKT 68
 QY 41 -----KGPQYPTW-----NAVLGWSLD--G 58
 DB 69 KLGSVGGQTDISIDYINPCVSSSGTTPCQEDSYGNWCKGKMGACNSQGIAYWSTD LFG 128
 QY 59 TSANPGDTFTLNM-----PCVFKYTTSQ-----TSVD 85
 DB 129 FYTPTPNV-TLEMTGYFLPQGTGYTFKATVDSDAISLVGGATAFNCCAQOQPITSTN 187
 QY 86 LTADGVK-----YATCPYSGEEFTFTSLCTVNDALKSSTKAFGT---VTLTP 131
 DB 188 FTIDGKPGGSLPPNIEGTVMYAGYYPM-----KVVYSNAVSGMTLPISVTLTP 238
 QY 132 IAFNVGGTGSSTDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFKSTV-DPSAYLYASRV 190
 DB 239 -----DGTTVSDDFEG-----YVVSFDD-----DLQSNCVDPDFSNYA-VSTTT 277
 QY 191 PSLNKVTTLFVAPQCENGYSCTGMGFSSSNGVDAIDCSNIHIGITGLNDWNPVSESF 250
 DB 278 TTTEPTGTGTSTSTEMTTVTGTNGVPTDFTVIRTPPTASTIITTEPNSTFTSTST 337
 QY 251 SYTKTCTSGIQIKYQNPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRQSKPFTL 310
 DB 338 ELTTVTGTNGVRDEII-----VIRTPPTATTAITTEPNSTFTSTSTSTEL----- 384

Db 50 DSSTYNAAYMAY---GYASKT-----KLGSGVGQDIDSIDYNIPCV-----SSGTPPCPO 98
QY 84 VDLTAD-GVK-YATCO-----FYSGEEFTTFLCTCTVNDALKSSIKAFGTWLPPIA-- 133
Db 99 EDLYGNWCKGIGACSNPIIAYWSDTLFGFYTPT-----NVTLEMTGY 143
QY 134 FNVGGTGSSTDLSDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLVARSWPML 193
Db 144 FLPPQTGSVT-----FKFATVDDSAI----- 165
QY 194 NKVTLFVAPOCENGYSGTMGFSNGDVAIDC-----SNHIGIKGLNDWN- 242
166 -----SVGSIATFCCAQEQPPTSTNFTI-----NGIKPWWG 198
QY 243 -YVSSSEFSYTKTCTSGNIQIKYQNVPA-GYRPF-----IDAYISATDVN- 286
Db 199 SPDNITGTYMYAGFYPMKIVYNAVANGTLPISVTLPGDGTVDGEGYVYTFDNNL 258
QY 287 --OYFLAYNDYTCAGSLQSPFLRWGKYNSDA-----GNGI-----VIVATP 331
Db 259 SQPNCIPDPSNVTVSTTTTTEP-----WTGFTSTSTEMTTVTGTNGVPTDETVIVIRT 314
QY 332 RTVTDSTTAVTLPNPVDKTKTIELQPIPTTIT-TSYVGVTTSYLTKTAPIGETAT 390
Db 315 PT-TASTIITTEPWNSTSTST-----ELTVTGTNGVPTDETVIVIRPTTATTA 366
QY 391 VIVDPYHHTTTTSEWGTITTTTTRTN--PTDSIDTVVQVPLPNPT-VSTTEYWSQS 447
Db 367 ITTTEPWNSTSTSTST-----TELTVGTNGLPDTE-TIIVIRPTTATTAMITQPNWDT 421
QY 448 FATT-----TTVTAPPG-GTD--TVIIRPPNHT--VTTTEYWSQSFAAT-----TTVTAPP 494
Db 422 FTSTSTELTVGTNGLPDTEIIVIRPTTATTAMITQPNWDTSTSTELTITVTGN 481
QY 495 G-GTDS-VIIRPPNPT-VTTTEYWSQSFAAT-----TTVTAPPG-GTDS--VIIRPP 542
Db 482 GLPTDETVIIRPTTATTAMITQPNWDTSTSTELTITVTGNLPTDETVIIRPTT 541
QY 543 NPT--VTTTEYWSQSFAAT-----TTVTAPPG-GTDS--VIIRPPNHT--VTTTEYWSQS 591
Db 542 TATTAMITQPNWDTST 601
QY 592 YAT-----TTVTAPPG-GTD--TVIIRPPNHT--VTTTEYWSQSFAAT-----TTVTGP- 637
Db 602 FTST 661
QY 638 --PSGTDTVIIRPPNPT-VTTTEYWSQSFAAT-----TTITAP-----PGETDVLIREPP 686
Db 662 GLPTDETVIIRPTTATTAMITQPNWDTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 721
QY 687 NHT--VTTTEYWSQSFAAT-----TTVTAP-----PGETDVLIREPPNHTV--TTTEYWSQS 735
Db 722 TATTAMITQPNWDTST 781
QY 736 YAT-----TTVTAPPG-GTD--TVIIRPPNPT--VTTTEYWSQSFAAT-----TTVTAPP 782
Db 782 FTST 841
QY 783 G--GTDTVIYESMS--SSKISTSSNDITSIIPS--FSRPHYVNSTTSDLSSTFSSSM 834
Db 842 GQPTDETVIIRPTTATTAMITQPNWDTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 901
QY 835 NPTTSSSDGM-----LLSSTTLVTESETTELICSDGKESRLSSSGSVITNPDNBSIV 891
Db 902 DTSSLVSSSVTSSSVTSSSVISSFSISSPVISSTTSSSILSESS-----K 948
QY 892 TSVTPASTMSDLSLSDGTSATSDNVSKSGSVTETSVITQIOTPNPLSSVTSLTQ 951
Db 949 SSVIPTSSSTSGSESETGAS 1004
QY 952 LSSIPVSESESKVTFSTNSGDNQSGTHDSQSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1011

Db 1005 TSSLPVPTTTKT-----SQDTLVTVTSCSHVCTESISSAIYSTAT 1046
QY 1012 PT---NTRQPTTLSTSTNSITDIT-TSQPTGDNQDNFTSSTNPVPTVATSLASEED 1067
Db 1047 VIVSGATTETVTCWCPISITETITKQTETTKQTKGTEQTTETTKQTTVVT---ISSCED 1103
QY 1068 --NKSCHESASTSLKPSGENSGLTSTTEIATTSPTPEAPSPAVSSGTDVTEPTDTR 1125
Db 1104 VCSKTASPAIVSTSTATNGVTVTETWCPITSTESKQQTTLVTVTSCSGVCSETT--- 1160
QY 1126 EOPTTLSTSTSKNSELVAT-----TQATNENGKSPSTDLTSSLTGTSTAST-----SA 1174
Db 1161 -SPAIVSTATATVNDVVTVSTWRPQTTNEQSVSSKMSATSETTINTGAAETTTSTGAA 1219
QY 1175 NSELVTSGSVT-----GGAVASANDQSHSTSV-----TNSNSIVSNTPTQTLTSSQVTS 1224
Db 1220 ETKTVTTSISIRFNAETQATSDVIGHSSSVSVSETGNTKSLTSSGLSTMSQOPRST 1279
QY 1225 -----SPSTNTFIASDYDGSIIQHSHTWLYGLITLLSLFI 1260
Db 1280 PASSVMYSGSTASLEISTYAGSANSLSLGLSVFTASLLAI 1321

RESULT 8
MUC2_HUMAN
ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN MUC2 OR SMUC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -!- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC AGENTS AT MUCOSAL SURFACES.
CC -!- SUBUNIT: MULTIMERIC.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND

CC VARIES AMONG DIFFERENT ALLELES.
 CC -!- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
 CC OF SILKWORM HEMOCYTIN.
 CC -!- SIMILARITY: CONTAINS 2 VWFC DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE (CTCK) DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: L21998; AAB95295.1; -;
 DR EMBL: M74027; AAA59875.1; -;
 DR EMBL: M94131; AAA59163.1; -;
 DR EMBL: M94132; AAA59164.1; -;
 DR Genew; HGNC:7512; MUC2.
 DR MIM; 158370; -;
 DR InterPro; IPR000359; Cys_knot.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002400; GF_cysknot.
 DR InterPro; IPR002919; TIL_Cysrich.
 DR InterPro; IPR001007; VWF_C.
 DR InterPro; IPR001846; VWF_D.
 DR Pfam; PF00007; Cys_knot; 1.
 DR Pfam; PF00093; vwc; 1.
 DR Pfam; PF00094; vwd; 4.
 DR Pfam; PF01826; TIL; 1.
 DR PRINTS; PR00438; GFCYSKNOT.
 DR SMART; SM00214; VWC; 2.
 DR SMART; SM00216; VWD; 4.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE; PS01185; CTCK_1; 1.
 DR PROSITE; PS01225; CTCK_2; 1.
 DR PROSITE; PS01208; VWFC; 2.
 KW Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 5179 MUCIN 2.
 FT DOMAIN 1401 1747 APPROXIMATE REPEATS.
 FT REPEAT 1401 1416 1.
 FT REPEAT 1417 1432 2.
 FT REPEAT 1433 1448 3.
 FT REPEAT 1449 1464 4.
 FT REPEAT 1465 1471 5.
 FT REPEAT 1472 1478 6.
 FT REPEAT 1479 1494 7A.
 FT REPEAT 1495 1517 7B.
 FT REPEAT 1518 1533 8A.
 FT REPEAT 1534 1556 8B.
 FT REPEAT 1557 1572 9A.
 FT REPEAT 1573 1596 9B.
 FT REPEAT 1597 1612 10A.
 FT REPEAT 1613 1635 10B.
 FT REPEAT 1636 1651 11A.
 FT REPEAT 1652 1675 11B.
 FT REPEAT 1676 1683 12.
 FT REPEAT 1684 1699 13.
 FT REPEAT 1700 1715 14.
 FT REPEAT 1716 1731 15.
 FT REPEAT 1732 1747 16.
 FT DOMAIN 4815 4886 VWFC 1.
 FT DOMAIN 4924 4991 VWFC 2.
 FT DOMAIN 5075 5160 CTCK.
 FT DISULFID 5075 5122 BY SIMILARITY.
 FT DISULFID 5089 5136 BY SIMILARITY.
 FT DISULFID 5098 5152 BY SIMILARITY.
 FT DISULFID 5102 5154 BY SIMILARITY.
 FT DISULFID 5159 5159 BY SIMILARITY.
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 770 770 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 894 894 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1154 1154 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1215 1215 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1230 1230 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1787 1787 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1820 1820 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4339 4339 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4351 4351 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4362 4362 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4373 4373 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4422 4422 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4438 4438 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4502 4502 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4616 4616 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4627 4627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4752 4752 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4787 4787 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4881 4881 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4888 4888 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4955 4955 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4970 4970 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5019 5019 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5038 5038 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5069 5069 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1351 1351 H -> L (IN REF. 3).
 FT CONFLICT 1412 1412 T -> S (IN REF. 3).
 FT CONFLICT 1449 1449 L -> P (IN REF. 3).
 FT CONFLICT 1504 1504 M -> T (IN REF. 3).
 FT CONFLICT 4192 4192 G -> S (IN REF. 2).
 SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;
 Query Match 11.1%; Score 721.5; DB 1; Length 5179;
 Best Local Similarity 25.3%; Pred. No. 1.7e-23;
 Matches 335; Conservative 158; Mismatches 554; Indels 279; Gaps 54;
 QY 59 TSANPGDFTLNMPCVFKYTTTSQTSVDLTADGVKATCFYSGEETFTSLCTVNDAL 118
 DB 1647 TTTTTPPTTTPSP-----TTTTTPPTTTPPTT-----TPSPITTTSPPT 1693
 QY 119 KSSIKAFGTVTLPIAFNVGGTGSSTDLSDSKFTAGTNTVTNDGDKDISIDVEFEKSTV 178
 DB 1694 TTTTTPPTTTPSSPITTTTTPSSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1747
 QY 179 DP---SAYLYASRVMPSLNKVT-----TLFVAPQCE-NGY-TSGPMGFSSNGDYA 224
 DB 1748 PTTTSSSLTTLPLPPSITPTTFSPFTTPTTTPCPLCNMTGWLDSGKPNFKPGDTE 1807
 QY 225 IDCSNIHIGITKGLNDWNPVSSEFSY-----TKTC-TSGIIGIKQN-VPAGY 272
 DB 1808 L-----IGDVCG-PGWAANISCRATPDVPIGOLGQTVWCDVSLGICKEDQKPGGV 1860
 QY 273 RPFIDAYISATDVNQVTLAYTNDYTCAGSRLOSKPFTLRWTGKKNDSNGIVIVATTR 332
 DB 1861 IPM-----AFCLNVEINVQCC--ECVQPTTMTTNTTENTPTTPTTPTTPTT 1907
 QY 333 TVDSTTAVTTLFPNPSVDKTKTIELQPIPTTTTTSYGVVTSYLTATAPIG-ETATV 391
 DB 1908 TPTPTGTGTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1967
 QY 392 IVDVPYHTTTTPTTSEWGT-----TTTTT-----TRNPPDSIDTVVQVP 432
 DB 1968 ---TTTTTTTPTPTPTPTPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2024
 QY 433 LPNPTVSTTEYWSQSFAITTTTAPPGGTDTVIIIEPPNHTVTTEYWSQSFAITTTTVA 492
 DB 2025 TPTPTGTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2072
 QY 493 PPGGD-----SVIIEPPNPTVTTEYWSQSFAITTTTPTTPTTPTTPTTPTTPTTPTT 535
 DB 2073 TPTGTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2132

QY 536 VIIEPPNPVTITTEYSQSYATTTTAP---PGTDSVILIREPPNHVTTTEYSQSY 592
 Db 2133 TTTTIVPTPTGTQTPPTTITTTTTPPTPTGTQ---PTTPTPTTT 2180
 QY 593 ATTTTAPPGGTGVILIREPPNHVTTTEYSQSFATTTTGTGSGTD-----T 643
 Db 2181 -TTTPTPTPTGTQ---PTTPTPTTT-----TTTPTPTPTGTQPTTPTTT 2226
 QY 644 VIIEPPNPVTITTEYSQSYATTTTAPGCTDVLIREPPNHVTTTEYSQSYATT 703
 Db 2227 TTTTPTPTGTQTP-TTTPTTITTTTPTPTGT-----QTPTTT-----PTTT 2272
 QY 704 TTTTAPPGETDVLIREPPNHVTTTEYSQSYATTTTAPPCTD-----T VII 754
 Db 2273 TTTTPTPTGT---QTPTPTPTTT-----TTTPTPTPTGTQTPPTTPTTTT 2321
 QY 755 REPPNPVTITTEYSQSFATTTTAPPGGTDTVIIYESMSSSKISTSSNITSLPFS 814
 Db 2322 TPTPTPTGTQTP-TTTPTTITTTTPTPTGT---QTPTPTTTT-----TTTPTPT 2372
 QY 815 RPHVNSVTSLSLFPSSSMPTPSSISDGLMLSSLTTLVTESETTELICSDGKCSRLS 874
 Db 2373 PTGTPTPTPTTT---TTTPTPTPTPTGTPTPTPTPTPTPTPTPTPTPTPT 2423
 QY 875 SSSGIVTNPDSNESSIVTSTPTASTMSDLSLSDIGISATS--SDNVSKSGSVTTTSV 932
 Db 2424 -----TPTPTPTTTTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 2476
 QY 933 TTIQT-----TPNPLSSVTSLSLTPSSVSESESKVT---FTSNGDNQSGTHDS 980
 Db 2477 TTTTPT 2536
 QY 981 QSTTEIEIVTSSKVLPT-----PVSSNPT-DLTSEPTNTREQPTTLSTNS 1028
 Db 2537 TQPTPTPTTTTPT 2596
 QY 1029 ITEDITSQPTGNDNTSSNPVPTVAIST-LASASEDNKSGSHESASLSKPSNGEN 1087
 Db 2597 VT---PTPTPT---GTQPTPTPTTTTPTPTPTPTPTPTPTPTPTPTPTPTPT 2650
 QY 1088 SGLTSTIEIATSPTEAPSPA-----VSSGDTVTTEPTDREQ----- 1127
 Db 2651 GTQPTPTPTTTTPT 2710
 QY 1128 -----PTTLSTKSTNSLVATQAT---NENGKSPSDDLSTLTGTSAASSELV 1179
 Db 2711 TTTTPTPTGTQTPPT 2770
 QY 1180 TSGSVTGGAVA-----SASNDQSHSTVNSNSVNTPTQTLTQOVTSSSPSTNTFTAST 1235
 Db 2771 TTTTPTTTTPT 1280
 QY 1236 YDGS GS 1241
 Db 2831 PTPTGT 2836

RESULT 9

FLO5_YEAST
 ID FLO5_YEAST STANDARD; PRT; 1075 AA.
 AC P38894;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DE 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Flocculation protein FLO5 precursor (Flocculin 5).
 GN FLO5 OR YHR21W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID:4932;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-S288C / AB972;
 RX MEDLINE-94378003; PubMed-8091229;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 Du Z., Favello A., Fulton C., Gattung S., Geisel C., Kirsten J.,
 Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
 Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
 Nhan M., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K.,
 Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 Vaudin M.;
 "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 VIII.";
 RL Science 265:2077-2082(1994).
 RN [2]
 RP REVIEW.
 RX MEDLINE-96076625; PubMed-7502576;
 RA Teunissen A.W., Steensma H.Y.;
 RT "Review: the dominant flocculation genes of Saccharomyces cerevisiae
 constitute a new subtelomeric gene family.";
 RL yeast 11:1001-1013(1995).
 CC -!- FUNCTION: MAY BE DIRECTLY INVOLVED IN THE FLOCCULATION PROCESS.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (Potential).
 CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED (PROBABLE).
 CC -!- SIMILARITY: BELONGS TO THE FLOCCULIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/annouce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U00029; AAB69731.1; -;
 DR PIR: S48992; S48992.
 DR SGD: S0001254; FLO5.
 DR InterPro: IPR001389; Flocculin.
 DR Pfam: PF00624; Flocculin; 8.
 KW Glycoprotein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 ? FLOCCULATION PROTEIN FLO5.
 FT PROPEP ? 1075 REMOVED IN MATURE FORM (POTENTIAL).
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 203 203 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 262 262 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 663 663 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 749 749 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT SEQUENCE 1075 AA; 111981 MW; D151B370B60C8D9F CRC64;
 SQ
 Query Match 10.9%; Score 711; DB 1; Length 1075;
 Best Local Similarity 25.9%; Pred. No. 8.3e-24;
 Matches 324; Conservative 192; Mismatches 393; Indels 342; Gaps 66;
 QY 153 AGTNTVTFNDGDKDISIDVEF-----EKSTVDPAYL---YASRV----- 189
 Db 23 SGATEACLPAGQRKSGNMNIFYOYSLKSDSYSTYNAAYMAYGASKTKLSVSGGTQDISD 82
 QY 190 --MPSLNKVTTLFVAPQCNGYTS-GTMGPFSSNGDVAIDCSNTHICITGLNDWNPVS 246
 Db 83 YNIPCVSSSGT-FPCFQ-EDSYNGWCKGMGA-----CSN-----SOGIAYW---S 123
 QY 247 SESFSYTKTSTNGIQKYNQVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSR---L 303
 Db 124 TDLFGFYTPTNTVLTGMTGYFLPPQGTSTFFSATVDDSAILSVGSGIAFECAQBPPI 183
 QY 304 QSKPFTLRWTGYNKSDAGSNGIVIVATT-----RTVDTSTAVTTLTPNPSVDKT 353
 Db 184 TSTNFTI--NGIKPMD-GSLPDNITGTVMYAGYYPLKVVYNAVSWGLPLI--SVE-- 236
 QY 354 KTEILQPIPT-TTITTSYGVGTVTSLTKTAPIGETATVIVDPYHTT---TTVTSEWTG 409
 Db 237 -----LPDGTTVSDNFEQYVYSP---DDLSQSNCTIPDPSIHTTITTTTPTWTG 285

```
QY 410 TIT-----TTTTRTNPTDSIDTVVQVPLPNPTVSTTEYWSQSFAITTTTAPPGGDT 463
D 410 TIT-----TTTTRTNPTDSIDTVVQVPLPNPTVSTTEYWSQSFAITTTTAPPGGDT 463
Db 286 TFTSTSTMTTITDNGOLDTETVIV---IKTPT-----TASTIT----- 322
QY 464 VIIREPPNHVTTTEYWSQSFAIT-----TWTAPPG-GTDS--VIIREPPNP--TWTTE 514
D 464 VIIREPPNHVTTTEYWSQSFAIT-----TWTAPPG-GTDS--VIIREPPNP--TWTTE 514
Db 323 -----TTTEPWTGTSTSTMTTITDNGOLDTETVIV---IKTPT-----TASTIT----- 322
QY 515 WQSFAIT-----TWTAPPG-GTDS--VIIREPPNP--TWTTEYWSQSFAITTTTAP 565
D 515 WQSFAIT-----TWTAPPG-GTDS--VIIREPPNP--TWTTEYWSQSFAITTTTAP 565
Db 372 PWTGTSTSTMTTITDNGOLDTETVIV---IKTPT-----TASTIT----- 322
QY 566 PGGDSVIIREPPNHVTTTEYWSQSFAITTTTAPPGTDTVIIREPPNH--TWTTEY 623
D 566 PGGDSVIIREPPNHVTTTEYWSQSFAITTTTAPPGTDTVIIREPPNH--TWTTEY 623
Db 432 ITGNG-----QPTDET-----IVRTTSEGLITTTTETP 462
QY 624 WQSFAIT-----TWTAPPG-GTDS--VIIREPPNP--TWTTEYWSQSFAITTTTAP 565
D 624 WQSFAIT-----TWTAPPG-GTDS--VIIREPPNP--TWTTEYWSQSFAITTTTAP 565
Db 463 WGTGTSTSTMTTITDNGOLDTETVIV---IKTPT-----TASTIT----- 322
QY 671 TAPPG-TD--TVLIIRPPNH--TWTTEYWSQSFAIT-----TWTAPPG-TD--TVLI 718
D 671 TAPPG-TD--TVLIIRPPNH--TWTTEYWSQSFAIT-----TWTAPPG-TD--TVLI 718
Db 523 TGTNGQPTDETIVIVIRTPTEGLITTTTETPWTGTSTSTMTTITDNGOLDTETVIV 582
QY 719 REPPNH--TWTTEYWSQSFAITTTTAPPGT-----DVTIIRPPNPVTTTETWSQ 770
D 719 REPPNH--TWTTEYWSQSFAITTTTAPPGT-----DVTIIRPPNPVTTTETWSQ 770
Db 583 RTPTSEGLITTTTETPWTGTSTSTMTTITDNGOLDTETVIV---IKTPT-----TASTIT----- 322
QY 771 SFATTTTVA-----PPGDTTV---IYESSMSSKISTSDNITSIIIPFSRPHY 818
D 771 SFATTTTVA-----PPGDTTV---IYESSMSSKISTSDNITSIIIPFSRPHY 818
Db 642 SGOITSSITSSRPIITPPYPSNGSVISSSVTSLSVTSFISSVISS----- 695
QY 819 VNSTSDLSSTESSMNT--PTSISDCMLSSITLVETSTETTELICSDKECSRLLSS 876
D 819 VNSTSDLSSTESSMNT--PTSISDCMLSSITLVETSTETTELICSDKECSRLLSS 876
Db 696 --STTTSIFESSSTSVITPSSSTGSSSEKTSASSSSSSSSISSSEPKSPNTNSSS 753
QY 877 SGIVTPDSNESSIIVTPVPASTMSDLSSTDGISATS-----SDNVKSGSVTTPET 930
D 877 SGIVTPDSNESSIIVTPVPASTMSDLSSTDGISATS-----SDNVKSGSVTTPET 930
Db 754 LPPVTSATGQET--ASSLPATTTKTS--EQTTLVTVTSCSHVCTEISSAIVSTAIVT 810
QY 931 --SVTTIOTTPNPLSSSVTSIQLSSIPSVSESKVFTTNGDNGQSTHDSQSTSTIE 988
D 931 --SVTTIOTTPNPLSSSVTSIQLSSIPSVSESKVFTTNGDNGQSTHDSQSTSTIE 988
Db 811 VSGVTEYTTWCPISTTET-----KQTKGTTETQKGTTEQT 847
QY 989 IVTTSSTKVLPPVSSNNTDLTSEPNTREQPTTLTSTNSITEDITTTSQPTGDCNDNTSS 1048
D 989 IVTTSSTKVLPPVSSNNTDLTSEPNTREQPTTLTSTNSITEDITTTSQPTGDCNDNTSS 1048
Db 848 TETTKQTTVV--TSSCEDICSKTAS-----PAIVSTSTATI-----NGVTTEY 890
QY 1049 TNPVPTVATSLASEEDNKGSGHESASTSLKPSMGNSGLITTSIEATTTSPTEAP- 1107
D 1049 TNPVPTVATSLASEEDNKGSGHESASTSLKPSMGNSGLITTSIEATTTSPTEAP- 1107
Db 891 TWCPISTES-----KQOTTLVTVTSCSGVCSETTSPA 925
QY 1108 --SPAVSGTDVTT-----EPTDREQPTTLTSTSKTNSLVAATQATNGKSPSPDL 1160
D 1108 --SPAVSGTDVTT-----EPTDREQPTTLTSTSKTNSLVAATQATNGKSPSPDL 1160
Db 926 IVSTATATVNDVTVVPTWRQTTNEQ-----SVSKMNS--ATSETTTNGAAETKTV 978
QY 1161 TSSLTGTSATSANSELVTSVSGVGGAVASANDQSHSTSV-----TNSNSIVSNTPT 1215
D 1161 TSSLTGTSATSANSELVTSVSGVGGAVASANDQSHSTSV-----TNSNSIVSNTPT 1215
Db 979 TSSLSRNFHAETQ-----TASATDVIGHSSVSVSEGTMTSMSTSSGLS 1023
QY 1216 TLSQVTS-----SPSTNTFIATYDGGSGIIIOHSTWLVGLTLLSLFI 1260
D 1216 TLSQVTS-----SPSTNTFIATYDGGSGIIIOHSTWLVGLTLLSLFI 1260
Db 1024 TMSQPRSTPASSVMWGSSTASLEISTYAGSANSLLAGSLVFASILLAI 1074
```

RESULT 10

```
AMYL YEAST STANDARD; PRT; 1367 AA.
ID AMYL YEAST
AC P08640; P08068;
DT 01-AUG-1988 (rel. 08, Created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
```

DE Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucohydrolase).

GN STAI OR STA2 OR MAL5 OR YIR019C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

ON NCBI_TaxID=4932;

OX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D., Churche C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A., Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M., Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D., Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V., Walsh S.V., Whitehead S.;

RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

RL [2]

RN SEQUENCE OF 1-242 AND 762-1331 FROM N.A.

RP MEDLINE=87194600; PubMed=3106330;

RX "Gene fusion is a possible mechanism underlying the evolution of STAI.";

RT J. Bacteriol. 169:2142-2149(1987).

RL [3]

RP SEQUENCE OF 1-31 FROM N.A.

RC STRAIN=SPX101-1C;

RX MEDLINE=89031230; PubMed=3141213;

RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;

RT "Similar short elements in the 5' regions of the STA2 and SGA genes from Saccharomyces cerevisiae.";

RL FEBS Lett. 239:179-184(1988).

CC -! CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose.

CC -! SIMILARITY: TO S.POMBE SPC215.13.

CC -! SIMILARITY: SOME, TO S.POMBE SPC285.13C.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; 238061; CAA86176.1; -

DR EMBL; M16164; AAA35014.1; -

DR EMBL; M16165; AAA35015.1; -

DR EMBL; X13857; CAA32069.1; -

DR PIR; B26877; B26877.

DR PIR; A26877; A26877.

DR PIR; S48478; S48478.

DR SGD; S0001458; MUC1.

KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein; Signal; Multigene family.

FT SIGNAL 1 21

FT CHAIN 22 1367

FT DOMAIN 210 1367

FT CARBOHYD 817 817

FT CARBOHYD 874 874

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;

Query Match 10.4%; Score 675.5; DB 1; Length 1367;

Best Local Similarity 26.2%; Pred. No. 3.6e-22;

Matches 378; Conservative 177; Mismatches 535; Indels 353; Gaps 65;

QY 1 MLQOFTLLFLYLISIASAKTITGVFDSFNSLWSNAANAFKPGVPTWNAVGLWGLDCTS 60

Db 1 MORPELLAYLVLSL-----FNSAL-----GFTALVPRG--SSESTS 36

QY 61 AN-----PGDTFTLNM--PCVFKYTTTSQTSVDLTADGVKATCYFSGSEFTFTSLTC 112

```
Db 37 CNSIVNGCNLFNHNHMOQNTQYTLDTVTSVWQDNT-YQITIHVKGKE----- 86
QY 113 TVNDALKS--SIKAFCTVTLPLAFNVGGTSGSTLDESKC-ETAGTNTVTEN-DGDKDI- 167
Db 87 --NIDLKYLWSLKIIGVCPKGTQVQLGYNEWTYLLIDNPTDFTA-----TFEYATQDVN 139
QY 168 -----STDVEFEKSTVDPSAYLXASRYMPSLKNKVTTLFLVAPQCENGYTSCTGMFSSN 220
Db 140 SCQVMPNFQIQE-----YLOGS-----AAQVASSMOWCTTSFDLST 177
QY 221 GDVAIDCSNIHIGITKGLDNWYPPVSSSEFSYTKTCSNGIOIKYQNVNPGAYRPRIDAYI 280
Db 178 G-----CNNDY--NOGHSQTFP-----GFYWNIDCDNN-----C 205
QY 281 SATDVNQYTLAYTNDYTCAGSLQSKPFTLRWTGYNKSDAGSNGIVIVATRTVTDSTA 340
Db 206 GGTKSSTTSSSESTTSSSESTT-----SSTSES-----TTSSTSESSTSS 254
QY 341 VTTLPNPSVDKTKTIEILOPIPTTTTTSYGVVTSYLTKTAPIGETATVIVDPYHTT 400
Db 255 STAPATPT-----TTSCTKEKPTPTTTS-----CTKEKP-----TPPHDT 292
QY 401 TTVTSWTCITITTTTRTN---PTDSIDIV-VVQVPLNPPTVSTTEYWSQ---SPATTIV 454
Db 293 TPCTKKTKTKCTCKTKTTPTPTPSSSTESSAPVPTPSSSTESSAPVPTSSSTESS 352
QY 455 TAP---PGCTDVIIRPPNHRVTTEYWSQFATTTVTAPPGGTDSVILIREPNTPTV 511
Db 353 SAPVPTPSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPT 461
QY 512 TTEYWSQFATTTVTAPPGGTDSVILIREPNTPTVTTTEYWSQFATTTT-VTAPPGGT 570
Db 402 TTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPT 461
QY 571 SVILIREPNTPTVTTTEYWSQFATTTVTAPPGGTDSVILIREPNTPTVTTTEYWSQFATTT 622
Db 462 SSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPT 521
QY 623 YWS-----QSEFATTTVTCPGGTDTVILIREP-PNPTVTTEYWSQFATTTTITAPGCE 676
Db 522 SSSAPAPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESS 575
QY 677 TDTVILIREPNTPTVTTTEYWSQFATTTVTAPPGGTDSVILIREPNTPTVTTTEYWSQFATTT 730
Db 576 SSSAPV---PTPSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESS 632
QY 731 YWS-----QSYATTTVTAP---PGCTDVIIRPPNHRVTTEYWSQFATTTTITAPGCE 781
Db 633 SSSAPAPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPT 690
QY 782 PGCTDVIIRPPNHRVTTEYWSQFATTTVTAPPGGTDSVILIREPNTPTVTTTEYWSQFATTTTITAPGCE 825
Db 691 SSSAPAPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPT 750
QY 826 ----LSTFESSMNTPTSSDGLLSTTLVTESETTTLICSDGKECSRLSSSGIVIN 882
Db 751 PVTSSSTESSAPVPTP-SSSTESSAPVPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESSAPVPT 809
QY 883 PDSNESSIVTIVPASTMSDSL-SSTDGISATSDNWSKGVSTTETSVTIOTTPNP 941
Db 810 PTPSSSNTTAPSTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESSAPVPTP 869
QY 942 LSSSVTSLTQLSIFPSVSES-----KVTFSTNGDNQSGTHDSQSTSEIEIV----- 990
Db 870 SSSSNTTAPSTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESS 929
QY 991 ----TTSSTKVLPPVSSNFDLSTSEPTN-----TREOPTTLST--TNSNITEDITSP 1038
Db 930 SVTTPSTTITTTVCSTGTNSAGETSGCSPKTVTTTPTTTTTSVTSSSTTTTITTVCS 989
QY 1039 TGDN--GDNTSNTNPVPTVATSLASASEDNKSGHESA-----ST 1078
Db 990 TGTNSAGETSGCSP-KTITTTVPCSTSPSEASESTTSPPTPTVTTVVSTTPTVTTTST 1048
```

1079 SLKPSMGNSGLTTS--TEIEATTSPTTEAPSPAV-----SSGTD----- 1116
1049 STKPG-GE---ITTFVTKNIPTTYLTITAPSPSTVTNFTPTTITTTVCSTGNSAGE 1104
1117 -----VTEPTDT-----REOPTTLSTSKTNSELVATQATNENGKSPSTDT 1161
1105 TTSKSPKTVTTVPCSTGTGEYTTTEATTLVTAFTTT-VVTTESSTGNSAGKTTTGYT 1163
1162 SLLTGTSTASTANSSELVTSG-----SVTGCVASASNDQSHSTSVTNSNSIVN----- 1211
1164 TKSVPVTVYTTLAPSAVPTPATNAVPTTITTECSAATNAAGETTSVCSAKTIVSSASAG 1223
1212 -----TPQPT---LSQQVTSSTSPSTN-----TFIASTYDGSISIIQ 1244
1224 ENTAPSAVPTVTAIPTVITTESSVGTNSAGETTTGYTKSIPTTYITLLIPGNGAKN 1283
1245 HST 1247
1284 YET 1286

RESULT 11
DAN4_YEAST
ID DAN4_YEAST STANDARD; PRT: 1161 AA.
AC P47179;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cell wall protein DAN4 precursor.
GN DAN4 OR YJR151C OR J2223.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Scarcez T.;
RN Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REGULATION.
RX MEDLINE=21113168; PubMed=11160904;
RA Cohen B.D., Sertliff O., Abramova N.E., Davies K.J., Lowry C.V.;
RT "Induction and repression of DAN1 and the family of anaerobic
RT mannoprotein genes in Saccharomyces cerevisiae occurs through a
RT complex array of regulatory sites";
RL Nucleic Acids Res. 29:799-808(2001).
CC -!- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE SRP1 / TIP1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 249651; CAA89684.1;
DR SGD: S0003912; DAN4.
DR InterPro: IPR000992; SRP1_TIP1.
DR Pfam: PF00660; SRP1_TIP1; 1.
DR PROSITE: PS00724; SRP1_TIP1; 1.
KW Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1146 CELL WALL PROTEIN DAN4.
FT PROPEP 1147 1161 REMOVED IN MATURE FORM (POTENTIAL).
FT LIPID 1146 1146 GPI-ANCHOR (POTENTIAL).
FT SEQUENCE 1161 AA; 118358 MW; 7954C15D69F0CA58 CRC64;

Query Match		8.9%; Score 577; DB 1; Length 1161;
Best Local Similarity		26.3%; Pred. No. 5e-18;
Matches		293; Conservative 131; Mismatches 427; Indels 264; Gaps 46;
QY	276 IDAYSATDVNOYTLAY-----TNDY--TCAGSRLOSKPFTLWHTGYKNSDAGSNGI	325
DB	36 IELAVVSDIRAHIFQYYSFRHHKHTETYPSEIAAAVFDYGDFTLRLTGISGDE-----	89
QY	326 VIVATRTVTDSTAVTLLPENPSVDKTKTIE-ILQPIPTTTTTSYVGVVSYLTKTAP	384
DB	90 -----VTRMTGVPWSTRI--KPAISSALSKGCIYTAIPTSTT-----TKSSTSTTP	138
QY	385 IGETATVIVDPVPHHTTTVTSEWGTITTTTTRTNPTDSIDVVQVPLPNTVSTTEYW	444
DB	139 -----TTTITS-----TTSTTSTPTTSTTST-----TPTTST---167	
QY	445 SOSFATTTVAPPGGTDVVIIRPNPHVTTTEWSQSFATTTVAPPGGTDSVLIIR	504
DB	168 STPTTSTSTTPTTSTTSTTPTTSTTPTTSTTPTTSTTPTTSTTPTTSTTPTT	227
QY	505 PNPVTVTTEWSQSFATTTVAPPGGTDSVLIIRPNPHVTTTEWSQSFATTTVTA	564
DB	228 STPTTSTTSTTSTA	272
QY	565 PPGGTDVVIIRPNPHVTTTEWSQSFATTTVAPPGGTDTVVIIRPNPHVTTTEYW	624
DB	273 PTTSTTS-----TTSTTS--TISTAPTSTSTSTSTSTSTSTSTSTSTSTSTSTA	317
QY	625 SOSFATTTVAPPGGTDVVIIRPNPHVTTTEWSQSFATTTVAPPGGTDSVLIIR	684
DB	318 STTFASLTTPAYSTASTDHT-----TSSVSTNAFTSATTTT-----SDTVISS	364
QY	685 PNPHTVTTEWSQSFATTTVAPPGGTDTVLIIRPNPHVTTTEWSQSFATTTVTA	734
DB	365 SFSQVTSSEPTVSEVSS--VEPTRSQVTSSEPTVSEVSSVEPTRSQVTSSEPT	418
QY	735 SYATTTTV---TAPPGGTDVVIIRPNPHVTTTEWSQSFATTTVAPPGGTDTVTA	781
DB	419 SSAEPTTVEFTSSVEPTRSQVTSSEPTVSEVSSVEPTRSQVTSSEPTVSEFT	477
QY	782 PGGTDTVVIIRPNPHVTTTEWSQSFATTTVAPPGGTDTVLIIRPNPHVTTTEWSQ	834
DB	478 TSSEPTRSQVTSSEPTVSEVSSVEPTRSQVTSSEPTVSEVSSVEPTRSQVTSSE	537
QY	835 NPTSTSSDGLSLSTLTVE---SETTT-----ELICSDGRECSRLSSS	876
DB	538 TTTEPVSSFGSEFTSSEPTVSEVSSVEPTRSQVTSSEPTVSEVSSVEPTRSQV	592
QY	877 SGIVTNPDNESSIVTSTV-PTA---STMSDLSLSTDTGISATTS-----	915
DB	593 SEIPSSIEVLTSSGSSSVSEPTSLVGPSSDESISTSPESLSATSTTSVAVSSKA	652
QY	916 -SDNVSKSGVSTTSTVTTIQTTPN-PLSSSVTSLTQ-----LSSIPSVSESS	963
DB	653 RSTVSAKSDVGNSSSTQSTTFATPSTPLAVSTVVTSTSDSVSPNIPSEISSSPES	712
QY	964 KVTFTSGNDGNQSDHSDSTSEIEIVTSSKVLPPVYSSNNTDLTSEPTNT-----	1018
DB	713 AITST	769
QY	1019 PPTLST	1073
DB	770 PLLVSSRSN--CSDARSS-----NTISSGLFSTIENVRNATSTFTNLSTD-----	812
QY	1074 ESASTSLKPSMENSGLTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST	1131
DB	813 EIVITSCKSSCTNEIDSVLTKQVSTVETITSC-----SGGICTTLMSPVTTINAK	866
QY	1132 STTSKTNSELVATQATNENGKS---PSTDLTSSLTWTCTSTASTSANELVTS-----	1181
DB	867 LTTTETST--VETITTCGGVCSTLTVPVTTTSTSTSTSTSTSTSTSTSTSTSTST	924
QY	1182 -----GSVTGGAVASASN-----DQSHSTSVTNSNSIVSNTPTQ-----	1223

DB	925 TLETTITSCGGICTTLMSPVTTINAKANTLTTTSTVETTTTSCGGVCSTLTPVPTT	984
QY	1224 -SSPSTNTTIASTYDGSGLIIHQHSTWLYGLITLLS	1257
DB	985 ITSEATTTTATISCEDNEEDVASTKTELLTMTTIT	1019
RESULT 12		
YS89_CAEEL	STANDARD; PRT; 3178 AA.	
AC	Q09624; Q09625; Q0969D4;	
DT	01-NOV-1995 (Rel. 32, Created)	
DT	15-JUN-2002 (Rel. 41, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Hypothetical protein_ZK945.9 in chromosome II.	
GN	ZK945.9/ZK945.10.	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;	
OC	Rhabditidae; Peloderinae; Caenorhabditis.	
OX	NCBI_TaxID=6239;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Bristol N2;	
RA	Wilkinson-Sproat J.;	
RL	Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	REVISIONS.	
RA	Durbin R.;	
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.	
CC	- SUBCELLULAR LOCATION: Integral membrane protein (Potential).	
CC	- SIMILARITY: CONTAINS 1 GPS DOMAIN.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
EMBL; Z48544; CAB70192.1; -		
DR	EMBL; Z48582; CAB70192.1; JOINED.	
DR	EMBL; Z48582; CAB70201.1; -	
DR	EMBL; Z48544; CAB70201.1; JOINED.	
DR	WormPep; ZK945.9; CE25697.	
DR	InterPro; IPR002111; Cat_channel_TrpL.	
DR	InterPro; IPR001024; Lipoxigenase_LH2.	
DR	InterPro; IPR000636; M+channel_nlg.	
DR	InterPro; IPR000203; PKD_cys_rich.	
DR	Pfam; PF00520; ion.trans; 1.	
DR	Pfam; PF01477; PLAT; 1.	
DR	Pfam; PF01825; GPS; 1.	
DR	SMART; SM00303; GPS; 1.	
DR	SMART; SM00308; LH2; 1.	
KW	Hypothetical protein; Transmembrane.	
FT	DOMAIN 266 1196 SER/THR-RICH.	
FT	DOMAIN 1105 1241 GLY/SER-RICH.	
FT	DOMAIN 2071 2120 GPS.	
FT	TRANSMEM 13 30 POTENTIAL.	
FT	TRANSMEM 51 73 POTENTIAL.	
FT	TRANSMEM 2139 2161 POTENTIAL.	
FT	TRANSMEM 2348 2367 POTENTIAL.	
FT	TRANSMEM 2390 2412 POTENTIAL.	
FT	TRANSMEM 2451 2468 POTENTIAL.	
FT	TRANSMEM 2483 2505 POTENTIAL.	
FT	TRANSMEM 2567 2589 POTENTIAL.	
FT	TRANSMEM 2836 2858 POTENTIAL.	
FT	TRANSMEM 2939 2961 POTENTIAL.	
FT	TRANSMEM 2976 2998 POTENTIAL.	
FT	TRANSMEM 3038 3060 POTENTIAL.	
QY	SEQUENCE 3178 AA; 344726 MW; F8239436D03666CD CRC64;	

Query Match 8.5%; Score 553.5; DB 1; Length 3178;
Best Local Similarity 24.3%; Pred. No. 1.6e-16;
Matches 290; Conservative 154; Mismatches 401; Indels 349; Gaps 51;

QY	252	YTKCTNSGIIQKYNVAGYRPEIDAYISATDVNQYTLAYTNDYTCAGRLOSKPETLR	311
DB	194	YREKCSGEINEIARRMC-KRPYRSE--KSTAISDSQGVYDGOVLKVR--AKOFMR	248
QY	312	WTG-----YKNSDANGSIVIVATRTVDSTAVTILPFPNSVDKTKTEILLQIPITTT	366
DB	249	TSGSPFLRMKRKADGN-----TCDYIEISTSTPTPTTVTSIVT-----STTT	295
QY	367	ITTSYVGVTSYLTAPIGETATVIVDPVYHTTTTWISEGTHTTTTTTRNTPDIDSIT	426
DB	296	VPTSTVTAMSTSTPSTSTTI-----ESTSTFTS-----TASTSTSTSTQQSSS	346
QY	427	VVOVPLNPNTVSTTEYSQSFAITTT--VTAPCGDTVIIRPPNHVTITTEYSQSF	484
DB	347	TITSSP-SSTTLST-----SIPTTTTPEITSTLSSLPDNAICSYLDETTTSTFTTMTL	399
QY	485	ATTITTAPPGGTDVIIIREPNPVITTEYSQSFAITTTVAPPGDTSVIAREPPNP	544
DB	400	TSITT-----EPTSTTTTEVTSTSVTTT-----EPTTT	431
QY	545	TVTITEYSQSFAITTTTAPPGGTDSVIIIREPNHVTITTEYSQSFAITTTVAPPGG	604
DB	432	LITSTASTSTTESPTSTVITS-----ISPVSITS-----SSSSSTVITVP---	475
QY	605	TDTVIIREPNHVTITTEYSQSFAITTTVGP-----PGDGTVIIRPPNPVTITTE	658
DB	476	TTESTSTSPSTVITS-----TTAPTSTTGPSSSSTSPSTASSVSTASSTOSSTS	530
QY	659	YKSQSFAITTT--TITAPGETDVIIRPPNHVTITTEYSQS-----YAT	702
DB	531	--TQGSTTKSETTSSDGNPDFVE----KATTFYDSTSNVNLASGLGIQYQT	584
QY	703	TTVTAPPGETDVIIRE-----PPNHTVTTT-----	729
DB	585	SICTSPTSNNVSTTKDGACFTKSVMPRLGGTYPASTFVPGNGYFRATMTDDKKVY	644
QY	730	-----EWWSQYATTT-----TVTAPPGTDVIIRPPNP-----TVTITEYSQSFAIT	775
DB	645	YYANYVIOEYSSTIESSESSTASVASSTST---PSTPSSTLSTSTVTEPSTRSSDS	700
QY	776	TTVTAPPGETDVIIREYSSMSKSTSNDLTISIIPFSRPHYVNVTSLDLSFESSMN	835
DB	701	TTISA--GSTTT--LQESTTSEESTDSSSTTTISDTSTSSPSTADSTS--TLSVD	754
QY	836	TPTSISDGMLLSST-----TLVTESETTELICSDGKECSRLLS-----	875
DB	755	QDFILDGLSNWNETRHEDSNIVLPNTAITER--SQTFECRNVTPEPLIKEST	812
QY	876	-----SSGITVNP-----DSN	886
DB	813	CLMYSNTLVNATYSSNIPIQTETFLVGIGTYEFRIINNLTMTMQVYSHIFTLNVADST	872
QY	887	ESSIVTSTVPT--ASTMSDSLSTOGISATSDNYKSCVSTTETSITTITQPENPLSS	945
DB	873	STSEVITSTSGSSSEAISTSGIESTSLEASTDASODSSTSTSDSGTT-----SD	927
QY	946	VTSLTQLSSIPVSESESKVFITFSGDNQSGTHDSQSTST-----EIEIVTT	992
DB	928	SITIDSSNSTPSTDSGG--LSQTPDSSSSASDSMRITTVDPDASTETPYDFVENLTW	984
QY	993	SSTK-----VLPVWSS-----NTDLTSEPNTREQPTTLSTNSITEDITTSOPTG	1040
DB	985	NETVYISENPFYITPIPKNEPGALTMATCQRDNDSSQPFVLLKESCLTE-----FG	1037
QY	1041	DNGDNTSST--NPVPTVATSTLASASEDNKSGSHASSTSLKPSMGEN--SGLTSTSTEIE	1097
DB	1038	KNGASVASFNPMYTSFVPAT-----GYEFLINVTNRAGESASHIFTNMVNL	1086
QY	1098	ATTSTPEAPSVA--GTDT--TEPTDTRQPTTLSTSTNSSELVATTO	1146

[illegible]

QY 720 EPNPHVTTEYSQSYATTVTAPPGGTDTVIIRPPNPTVTTEYSQSFATTTT 779
 Db 347 EAPSTSSVSSISSTSSVSEAPLATSSVSSSEAPSTSSVSSSEAPSTSSVSS 406
 QY 780 APGCGTDTVIIESMSSKISSTNDITSIPSPRPHVNSTSDLSSTFSSMMTPS 839
 Db 407 EAPSTSSVSSSE-ISTKSSVSSSE-----VSSATSSVSSSEAPS--AISS 450
 QY 840 ISSDGLLSTTLVTESETTEL-ICSDGKESRLSSSGSVIWNPDN--ESSIVTSV 895
 Db 451 LASSRLFSKNTSVTLVATEASSVTSLSRPSSETLASNIIESLSGYNSTVSTTS 510
 QY 896 PTASTMSDLSSTDGISATS-----SDNVSKGVSVTETSTVT-----IOTPNPLSSS 945
 Db 511 AASSTLGSVSSNSRMATSKTSSSTSSDLKSSVIFGNSSTVTPSPASISLTSAPLPSV 570
 QY 946 VTSLTQSLSPVSESEKVTFTSNGDNGSGHDSOSTSTEIE-IVTTSKVLPPVSS 1004
 Db 571 WSDITS-----SEASS-----ISSNLASSAFSDNNSTIASASLIVTKNSVYSSIVSS 620
 QY 1005 NTDLTSEPNTREOPTTLSTNSITEDITTOPTGDNDSSTNPVPTVATSTLASAS 1064
 Db 621 IT--SSETYN-----ESNLATSTSLSNKATARSLS--NATSANVTGTFFSSMSSHT 672
 QY 1065 EEDNKGSHESASTSLKPMGENS--GLTSTTEIEATTS---PTPEAPSVSSSGTDTVT 1119
 Db 673 SVITPGFSTSSALAINSTVSSSLAGYSFSPESPSTTLVTSSEAPS--TVSSMT--TS 729
 QY 1120 EP-----TDTREOPTTLSTKNSLSELVATQNTENGKSPSTDLTSLTTGTSA---- 1170
 Db 730 APINNSTGARPSPASTITESTSISVPLASGDVTSLSLAHNITTSFAPSTSSAQLV 789
 QY 1171 --STANSSELVT-----SGSVTGA-----VASANDQSHSTSVTNSNSIVSN 1211
 Db 790 SKSTSSSLVPRIDRSGNSSTASRIATSLPNKTFVSSLSTSAHARINFNSTVIATA 849
 QY 1212 TPOTTLQQVSSSPNTFFIATSDGSGSIIOHSTWLYGLITLLS 1257
 Db 850 KQIETLSTVNCNPNPNITKV-----IVSRETTAIGTVTSCS 890

RESULT 15
 YK82_YEAST
 ID YK82_YEAST STANDARD; PRT; 1169 AA.
 AC P36170;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Hypothetical 122.2 kDa protein in SIR1 3' region precursor.
 GN YKR102W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gaillon L., Dufon B.;
 RL Submitted (Mar.1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: CELL WALL (BY SIMILARITY).
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY).
 CC -1- SIMILARITY: STRONG, TO YEAST PROTEIN FLO1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z38327; CAAB2182.1;
 DR PIR; S38181; YKR102W.
 DR SGD; S0001810; YKR102W.

KW Hypothetical protein; Glycoprotein; Repeat; Cell wall; Signal.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 1169 HYPOTHETICAL PROTEIN YKR102W.
 FT DOMAIN 274 609 SER/THR-RICH.
 FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 279 279 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 389 389 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 452 452 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 656 656 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 686 686 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 879 879 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1092 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1099 1099 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1169 AA; 122164 MW; 940C6DFB3569C669 CRC64;
 Query Match 7.7%; Score 501; DB 1; Length 1169;
 Best Local Similarity 23.1%; Pred. No. 9.2e-15;
 Matches 313; Conservative 208; Mismatches 528; Indels 308; Gaps 53;
 QY 6 TLLFLYLSTASAKTIT-----GVDSFNSLTWSNAANYAFKGGYPTWNAVLG 53
 Db 11 TGLFL-LSVANVALGTTEACLPAGEKKNGMTINFYQSLKDSSTYS--NPSYMAYGADA 67
 QY 54 WSLDGTSANPGDTFTLNMPCVKYTTTSQTSVDLTADGVKYATCOFYSGE-----EFTT 106
 Db 68 EKLGSVSGOTKLSIDYSIPCNASDTCACSD--DATEYSASQVVPVVRGVKLCSDNTT 124
 QY 107 FSTLT-----CTVDALKSSIKAFGTVTLPIAFNVGGTSGSSTDLSEKCFAGTAGNTVT 159
 Db 125 LSKTEKRENDCCDQGAAYWSS-DLFGFYTPTNIVEMTG-----YFLPPKGTGTYT 175
 QY 160 FNDGDKDIDVEFEKSTVDPSLAYLYASRMPSLNKVTTLFVAPQCENGYTSCTMGFSSS 219
 Db 176 FG-----FATVDDSAIL-----SV 189
 QY 220 NGDVAIDCSNIH---IGIT---KGLNDWN--YPVSSSEFSYTKTCTNSGIIQIKYON-VP 269
 Db 190 GGNVAFECCKQEQPPITSTDTFTINGIKPNADAPTIDKISTYWIAGYYPKIKIVISNAVS 249
 QY 270 AGYRPPIDAYISATDVNQ---YTLAYTND-----YTCAGSRLQSKPFTL 310
 Db 250 WGTLPVSVVLPDGTVEVDDDFEGYVDFSDNATGHCSPNPAEHARTCVSSATSS----- 304
 QY 311 RWTGKNSDAGSNGIVATRTVTSTTAVTLPNPSVDKTKTEILQIPPTTTTTS 370
 Db 305 -WSSSE-----VCTETETESTSYVTPYVTSSSSSSEVCTECTETESTSTSTP 352
 QY 371 YGVVTSYLTKT-APIGETATVIVDPYHTTTVTSEMTGTTITTT---TRTNPTDSIDT 426
 Db 353 YVTSSSSSEVCTECTETESTSYVTPYVSSSTAAANYTSSSSSEVCTECTETESTST 412
 QY 427 VVQVPLPNPTVSTTEYSQSFAATTTVAPPGGTDTVIIRPPNHTVTTEYSQSFAF 486
 Db 413 STPYV-----TSSWSSEVCTECTE-----TESTSYVTPYVSSSTAAANYTSSFS 459
 QY 487 TTTVTAPPGGTDSVIIRPPNPTVTTEYSQSFAATTTVAPPGGTDSVIIRPPNPTV 546
 Db 460 SSEVCTECTETES-----TSTPYVTSSSSSSEVCTECTETESTSYVTVSSS 510
 QY 547 TTEYSQSFAATTTVAPPGGTDSVIIRPPNHTVTTEYSQSFAATTTVAPPGGTDT 606
 Db 511 TAAANYTSSFSSEVCTECTETES-----TSTPYVTSSSWSSEVCTECTETE 561
 QY 607 TVIIRPPNHTVTTEYSQSFAATTTVTPGPGTDTVIIRPPNPTVTTEYSQSFA 666
 Db 562 STSYVTPYVSSSTAAANYTSSFSSEVCTECTETES-----TSTPYVTSSTGT 612
 QY 667 TTTITAPPGETDTVLIRPPNHTVTTEYSQSFAATTTVAPPGETDTVLIRPPNHTV 726
 Db 613 ATSFTASTNTMTSLVQ-----TDPTVSFSLSTVSEHTNAP-----TSSVESNASTFI 661

```

QY 727 TTEYWSQYATTTVTAPPGTDTVIIRPPNPVT-----TTEYWSQSFATTTVTAP 781
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 662 SSNKGSKSVTSIHITP-----MYPSNQVTSSVSVPTTSESESSASVTIL 713
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 782 PGCTDVIYVESSSKISTSSNDITSIPSF---SRPHVYNSTTDLSTFESSMNTPT 838
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 714 P-STITSEFKPSTMTKTVWSISSPTNLITSYDTKSDTVGSGTSSVLSISLPSY 772
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 839 SISSD---CMLSSSTTLVTESETTELICDCKECSRLSSSGIYNPDNNESSIVTST 894
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 773 SASSEQIFHSIVSSNGQALTFSSTKVSSSESESHRTSPPT-----SSBSGKSSG 825
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 895 VPTASTMSDLSSTDGISTSDNVKSGSVVTTETSVTTIQTTPNPLSSSVTSLQLSS 954
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 826 VEIESTSTSFSPHE--TSTASTVQISSQFVTPSPISVA---PRSTGLNSQIE--- 876
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 955 IPSVSESEKVTTSNGDNQSGTHDSQSTETIEIVTTSKVLPPVVSNTD----- 1007
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 877 -----STNSSKETMSSE-NSASVMPSSSATSPKTKVTSDET-----SSGFSRDRITVY 924
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1008 -LTSEPTNTREOPTTLSTT-----SNSITEDITTSQPTGNDGNDTSSNPNVPTVATST 1059
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 925 RMSETPSTNEQTTLTVSSCESNSCNTVSSAVVSTATTINGITTEYTWCPPLSATE- 983
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1060 LASASEEDNKGSHESASTSLKPSMGENSEGLTSTETETATTSPTAP---SPAVSSGTD 1116
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 984 LTTVSKLESE-----EKTTLITVTSCEGVCSETASPAIVSTATATVND 1027
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1117 VTTEPTDTRQPTT-LSTTSKTNSEL-----VATTQATNENGKSPSTDLTSSLT-T 1166
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1028 VTVVYSTWSPQATNKLAVSDIENSASKASFVSEAAETKISRNNNFVPTSGTSTETHT 1087
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1167 GTSASTSANSSELTSGSVTCGAVASASANDOSHSTSVTNSNI-VSNTPQTTLSQQVTSS 1225
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1088 TTTNSASESNDV-----SASEAVS-SKSVTNPNVLISVSOQPRGTPASSMIGS- 1134
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1226 PSTNTFIATYDGSIIQHSTWLYGLITL--LSLFI 1260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1135 -STASLEMSSYLG---IANH-----LLTNSGISIFI 1161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: June 11, 2003, 17:12:10
Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:12:16 ; Search time 298 Seconds
(without alignments)
2726.053 Million cell updates/sec

Title: US-09-715-876-8

Perfect score: 6495

Sequence: 1 MLQOFTLLFLYLSIASAKTI.....SIHQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_AA_Main.*

```
1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US097_COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US098_COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US099_COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US100_COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US101_COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US102_COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US60_COMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	6495	100.0	1260	21	US-09-715-876-8
2	6465	99.5	1260	27	US-60-385-568-211
3	4204.5	64.7	1119	27	US-60-385-568-213
4	4190.5	64.5	1191	16	US-09-248-796-16243
5	4190.5	64.5	1191	27	US-60-096-409-16243
6	4105	63.2	828	25	US-10-179-131-7791

7	4078.5	62.8	1047	27	US-60-385-568-218	Sequence 218, App
8	3951	60.8	1586	27	US-60-385-568-295	Sequence 295, App
9	3777.5	58.2	885	25	US-10-179-131-8318	Sequence 8318, App
10	3777.5	58.2	886	27	US-60-385-568-297	Sequence 297, App
11	3765.5	58.0	1756	27	US-60-385-568-322	Sequence 322, App
12	3739	57.6	1182	25	US-10-179-131-7998	Sequence 7998, App
13	3736.5	57.5	1581	25	US-10-179-131-10126	Sequence 10126, A
14	3716.5	57.2	1523	27	US-60-385-568-214	Sequence 214, A
15	3620	55.7	1593	27	US-60-385-568-323	Sequence 323, App
16	3586	55.2	1332	25	US-10-179-131-7938	Sequence 7938, App
17	3375	52.0	657	25	US-10-179-131-6795	Sequence 6795, App
18	3220	49.6	1060	16	US-09-248-796-14123	Sequence 14123, A
19	3220	49.6	1060	27	US-60-096-409-14123	Sequence 14123, A
20	3194	49.2	1419	27	US-60-385-568-215	Sequence 215, App
21	3067.5	47.2	952	27	US-60-385-568-319	Sequence 319, App
22	3067	47.2	894	27	US-60-385-568-287	Sequence 287, App
23	3015.5	46.4	819	25	US-10-179-131-5983	Sequence 5983, App
24	2937	45.2	1347	27	US-60-385-568-361	Sequence 361, App
25	2791.5	43.0	1443	27	US-60-385-568-216	Sequence 216, App
26	2779	42.8	780	16	US-09-248-796-16702	Sequence 16702, A
27	2779	42.8	780	27	US-60-096-409-16702	Sequence 16702, A
28	2623	40.4	775	25	US-10-179-131-8127	Sequence 8127, App
29	2582	39.8	1029	27	US-60-385-568-293	Sequence 293, App
30	2576.5	39.7	2297	27	US-60-385-568-217	Sequence 217, App
31	2560	39.4	1189	25	US-10-179-131-6661	Sequence 6661, App
32	2548.5	39.2	2000	27	US-60-385-568-409	Sequence 409, App
33	2544	39.2	1366	27	US-60-385-568-411	Sequence 411, App
34	2544	39.2	1384	25	US-10-179-131-9947	Sequence 9947, App
35	2320	35.7	909	25	US-10-179-131-6405	Sequence 6405, App
36	2297	35.4	646	16	US-09-248-796-16701	Sequence 16701, A
37	2297	35.4	646	27	US-60-096-409-16701	Sequence 16701, A
38	2273	35.0	812	16	US-09-248-796-16244	Sequence 16244, A
39	2273	35.0	812	27	US-60-096-409-16244	Sequence 16244, A
40	2271	35.0	467	27	US-60-385-568-288	Sequence 288, App
41	2243.5	34.5	895	25	US-10-179-131-5330	Sequence 5330, App
42	1994	30.7	511	16	US-09-248-796-14125	Sequence 14125, A
43	1994	30.7	511	27	US-60-096-409-14125	Sequence 14125, A
44	1986.5	30.6	468	27	US-60-385-568-212	Sequence 212, App
45	1973	30.4	451	25	US-10-179-131-5151	Sequence 5151, App

ALIGNMENTS

RESULT 1

US-09-715-876-8

; Sequence 8, Application US/09715876

; GENERAL INFORMATION:

; APPLICANT: Edwards Jr., John E., et al.

; TITLE OF INVENTION: Pharmaceutical Compositions and Methods to Vaccinate Against

; TITLE OF INVENTION: Candidiasis.

; FILE REFERENCE: 259/064

; CURRENT APPLICATION NUMBER: US/09/715,876

; CURRENT FILING DATE: 2000-11-18

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 8

; LENGTH: 1260

; TYPE: PRT

; ORGANISM: Candida albicans

; US-09-715-876-8

Query Match 100.0%; Score 6495; DB 21; Length 1260;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLQOFTLLFLYLSIASAKTITGVDFNSLTWNAAYAFKPGPYPTWNAVLSLDGTS	60
DB	1	MLQOFTLLFLYLSIASAKTITGVDFNSLTWNAAYAFKPGPYPTWNAVLSLDGTS	60
QY	61	ANPGDFTLNMPCVFYKTTTSQTSVDTADGVKYATCOFYSGEEFTTSTLTCTVNDALKS	120
DB	61	ANPGDFTLNMPCVFYKTTTSQTSVDTADGVKYATCOFYSGEEFTTSTLTCTVNDALKS	120

QY 121 SIKAFGTVTLPIAFNVGGTSGSTDLSDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
DB 121 SIKAFGTVTLPIAFNVGGTSGSTDLSDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
QY 181 SAYLYASRVMPSLNKVTTFLFVAPQCENGYTSCTMGFSSNGDVAIDCSNIHIGITKGLND 240
DB 181 SAYLYASRVMPSLNKVTTFLFVAPQCENGYTSCTMGFSSNGDVAIDCSNIHIGITKGLND 240
QY 241 WNPVSSSEFSYTKTCTSGNGIQIKYQNPAGYRPFIDAYISATDVNQYTLATNDYTCAG 300
DB 241 WNPVSSSEFSYTKTCTSGNGIQIKYQNPAGYRPFIDAYISATDVNQYTLATNDYTCAG 300
QY 301 SRLQSKPFTLRWGTGKNSDAGSNGIVIVATRTVTDSTTAVTTLFPNPSVDKTKTIELQ 360
DB 301 SRLQSKPFTLRWGTGKNSDAGSNGIVIVATRTVTDSTTAVTTLFPNPSVDKTKTIELQ 360
QY 361 PIPTTTITTSYGVVTSYLTKTAPIGETATVIVDPYHTTTVTSEWGTITTTTTRNP 420
DB 361 PIPTTTITTSYGVVTSYLTKTAPIGETATVIVDPYHTTTVTSEWGTITTTTTRNP 420
QY 421 TDSIDTVVQVPLPNTVSTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEY 480
DB 421 TDSIDTVVQVPLPNTVSTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEY 480
QY 481 SOSFATTTVTAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIR 540
DB 481 SOSFATTTVTAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIR 540
QY 541 PNPNTVTTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVT 600
DB 541 PNPNTVTTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVT 600
QY 601 PPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEY 660
DB 601 PPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEY 660
QY 661 SOSYATTTITAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIR 720
DB 661 SOSYATTTITAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIR 720
QY 721 PNPNTVTTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVT 780
DB 721 PNPNTVTTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVT 780
QY 781 PPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEY 840
DB 781 PPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEY 840
QY 841 SSDGMLLSSTLLVTESETTELICSDGKBCSRLSSSGIVTNPDSNESSIVTSTVPTAST 900
DB 841 SSDGMLLSSTLLVTESETTELICSDGKBCSRLSSSGIVTNPDSNESSIVTSTVPTAST 900
QY 901 MSDLSSTDGISATSSDNVSKGVSTTETSVTTTQTPNPLSSSVTSLTQLSSIPSVSE 960
DB 901 MSDLSSTDGISATSSDNVSKGVSTTETSVTTTQTPNPLSSSVTSLTQLSSIPSVSE 960
QY 961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVYVSSNTDLTSEPTNREOPT 1020
DB 961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVYVSSNTDLTSEPTNREOPT 1020
QY 1021 TLTSTNSITEDIQTSQPTGONGDNTSSTNPVPTVATSTLASASEEDNKGSGHESASTSL 1080
DB 1021 TLTSTNSITEDIQTSQPTGONGDNTSSTNPVPTVATSTLASASEEDNKGSGHESASTSL 1080
QY 1081 KPSMGENSEGLTSTTELEATTTSPTEAPSPAVSSGTDVTEPTDTRREQPTLLSTTKTNS 1140
DB 1081 KPSMGENSEGLTSTTELEATTTSPTEAPSPAVSSGTDVTEPTDTRREQPTLLSTTKTNS 1140
QY 1141 LVATTQATNENGGKSPSTDLTSSLATGTASSTANSELVTSVSGVTGGAVASANDQSHST 1200
DB 1141 LVATTQATNENGGKSPSTDLTSSLATGTASSTANSELVTSVSGVTGGAVASANDQSHST 1200

QY 1201 SVTNSNSIVSNTPQTTLTSSQVTSSTSPSTNTFIASVYDGSIIQIHSWTWLYGLITLLSLFI 1260
DB 1201 SVTNSNSIVSNTPQTTLTSSQVTSSTSPSTNTFIASVYDGSIIQIHSWTWLYGLITLLSLFI 1260
RESULT 2
US-60-385-568-211
; Sequence 211, Application US/60385568
; GENERAL INFORMATION:
; APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Glandong Zeng
; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and their Use
; FILE REFERENCE: 032796-134
; CURRENT APPLICATION NUMBER: US/60/385,568
; PRIOR FILING DATE: 2002-06-05
; NUMBER OF SEQ ID NOS: 2002
; SEQ ID NO 211
; LENGTH: 1260
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-385-568-211
Query Match 99.5%; Score 6465; DB 27; Length 1260;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1255; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 MLQOFTLLFLYLISASAKTITGVDFSNLSLWNSAANAFKPGYPTWNAVLCWSLDGTS 60
DB 1 MLQOFTLLFLYLISASAKTITGVDFSNLSLWNSAANAFKPGYPTWNAVLCWSLDGTS 60
QY 61 ANPGDTFTLNMPCVEKYTTTSTQSVDLTADGVKYATCQFYSGEFTTFTSLTCTVNDALKS 120
DB 61 ANPGDTFTLNMPCVEKYTTTSTQSVDLTADGVKYATCQFYSGEFTTFTSLTCTVNDALKS 120
QY 121 SIKAFGTVTLPIAFNVGGTSGSTDLSDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
DB 121 SIKAFGTVTLPIAFNVGGTSGSTDLSDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
QY 181 SAYLYASRVMPSLNKVTTFLFVAPQCENGYTSCTMGFSSNGDVAIDCSNIHIGITKGLND 240
DB 181 SAYLYASRVMPSLNKVTTFLFVAPQCENGYTSCTMGFSSNGDVAIDCSNIHIGITKGLND 240
QY 241 WNPVSSSEFSYTKTCTSGNGIQIKYQNPAGYRPFIDAYISATDVNQYTLATNDYTCAG 300
DB 241 WNPVSSSEFSYTKTCTSGNGIQIKYQNPAGYRPFIDAYISATDVNQYTLATNDYTCAG 300
QY 301 SRLQSKPFTLRWGTGKNSDAGSNGIVIVATRTVTDSTTAVTTLFPNPSVDKTKTIELQ 360
DB 301 SRLQSKPFTLRWGTGKNSDAGSNGIVIVATRTVTDSTTAVTTLFPNPSVDKTKTIELQ 360
QY 361 PIPTTTITTSYGVVTSYLTKTAPIGETATVIVDPYHTTTVTSEWGTITTTTTRNP 420
DB 361 PIPTTTITTSYGVVTSYLTKTAPIGETATVIVDPYHTTTVTSEWGTITTTTTRNP 420
QY 421 TDSIDTVVQVPLPNTVSTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEY 480
DB 421 TDSIDTVVQVPLPNTVSTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEY 480
QY 481 SOSFATTTVTAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIR 540
DB 481 SOSFATTTVTAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIR 540
QY 541 PNPNTVTTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVT 600
DB 541 PNPNTVTTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVT 600
QY 601 PPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEY 660
DB 601 PPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIRPPNHTVTTTEY 660
QY 661 SOSYATTTITAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIR 720
DB 661 SOSYATTTITAPPGGTDVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGTDVVIIR 720


```

RESULT 4
US-09-248-796-16243
; Sequence 16243, Application US/09248796
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 16243
; LENGTH: 1191
; TYPE: PRF
; ORGANISM: Candida albicans
US-09-248-796-16243

Query Match      64.5%; Score 4190.5; DB 16; Length 1191;
Best Local Similarity 68.1%; Pred. No. 1.6e-296;
Matches 863; Conservative 89; Mismatches 195; Indels 121; Gaps 19;

QY 1 MLOQFTLLFLYLISIAKTTTGVDSPNSLTWSNAANYAFKPGYPTWNAVLGWSLDGTS 60
DB 3 MLOQFTLLFLYLISIAKTTTGVDSPNSLTWSNAANYAFKPGYPTWNAVLGWSLDGTS 62
QY 61 ANPGDTFTLNMPCVFYKTTTSQTSVDLTADGKVATCFYSGEEFTTSTLCTVNDALKS 120
DB 63 ANPGDTFTLNMPCVFYKTTTSQTSVDLTADGKVATCFYSGEEFTTSTLCTVNDALKS 122
QY 121 SIKAFGTVTLPIAFNVGGTSSDLEDSKCFCTAGTNTVTENDGDKDISIDVEFEKSTVDP 180
DB 123 SIKAFGTVTLPIAFNVGGTSSDLEDSKCFCTAGTNTVTENDGDKDISIDVEFEKSTVDP 182
QY 181 SAYLYASRVMPSLNKKVTTLEVAPOCENGYSGTMGFSSNGDVAIDCSNIHIGITKGLND 240
DB 183 SGVLYASRVMPSLNKKVTTLEVAPOCENGYSGTMGFSSNGDVAIDCSNIHIGITKGLND 242
QY 241 WNPVSSSESYTKTCSNGIQIKYQNVAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
DB 243 WNPVSSSESYTKTCSNGIQIKYQNVAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 302
QY 301 SRLQSKFETLRWTKYKNSDAGSNGIVATRTVDTSTTAVTTLFNPSPVDKTKTIELQ 360
DB 303 SRLQSKFETLRWTKYKNSDAGSNGIVATRTVDTSTTAVTTLFNPSPVDKTKTIELQ 362
QY 361 PIPTTTTTSYGVGTTSYLTKTAPIGETATVVDVPHYHTTTTTSWTGTTTTRTNP 420
DB 363 PIPTTTTTSYGVGTTSYLTKTAPIGETATVVDVPHYHTTTTTSWTGTTTTRTNP 422
QY 421 TDSIDTIVVQVPLPNPTVSTTEYSQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYW 480
DB 423 TDYIDTIVVQVPLPNPTVSTTEYSQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYW 482
QY 481 SQSFATTTTAVPPGTDVVIIRPPNPTVTTTEYSQSFATTTTAVPPGTDVVIIR 540
DB 483 SQSFATTTTAVPPGTDVVIIRPPNPTVTTTEYSQSFATTTTAVPPGTDVVIIR 542
QY 541 PPNTPTVTTTEYSQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYSQSFATTTT 600
DB 543 PPNTPTVTTTEYSQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYSQSFATTTT 602
QY 601 PPNGTDVVIIRPPNHTVTTTEYSQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYW 660
DB 603 PPNGTDVVIIRPPNHTVTTTEYSQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYW 662
QY 661 SQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYSQSFATTTTAVPPGTDVVIIR 720
DB 663 SQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYSQSFATTTTAVPPGTDVVIIR 722
QY 721 PPNTPTVTTTEYSQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYSQSFATTTT 780
DB 723 PPNTPTVTTTEYSQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYSQSFATTTT 782

```

```

DB 723 PPNTPTVTTTEYSQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYSQSFATTTT 782
QY 781 PPNGTDVVIIRPPNHTVTTTEYSQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYW 840
DB 783 PPNGTDVVIIRPPNHTVTTTEYSQSFATTTTAVPPGTDVVIIRPPNHTVTTTEYW 816
QY 841 SS--DGMLLSSTTLVTE---SETTELICSDGKESRSLSSSGIVGNPDSSNESSIVTST 894
DB 817 TGPPGG--TDTVIIRPPNPTVTTTEY-----WSQSYATTT-----TDT 853
QY 895 VPTASTMSDLSSTDGISATSSDNVSKSGVSVTETSVTTIQTTPNPLSSSVTSLTQLSS 954
DB 854 APFGGTATVIIRPPNPTVTTTEYSQSFATTTT-----GPPGTDVVIIRPP 904
QY 955 IPSVSESESKVTFTSNGDNQSGTHDSQSTSEIEIVTTSKVLPPVVSNTDLTSPTN 1014
DB 905 SPVITTE-----YWSQSYAT-----TDTVAPPGTDVVIIRPPN 941
QY 1015 TREQPTTLST--SNSITEDITTSQPTGDNDTSSNPNVPTVATSTLASAEDNKSGS 1072
DB 942 ----PTVTTTEYSQSFATTTTVPGGTDVVIIRPPSPVTTT-----EYWSQS 989
QY 1073 HESASTSLKPMGENSEGLTST----EIEATTSPTPEAPVSSGTDVTEP--TDT-- 1124
DB 990 YATTTVTAPPGG-----TATVIIRPPNPTVTTTEYSQSFATTTTVPGGTDV 1043
QY 1125 -REQPTTLSTTSNSELVATTOATNENGKSPSTDLTSSLTGTSTASTSANSSELVT-- 1180
DB 1044 IRPPNPTVTTTEYSQSFATTTTITAPPGTDVVIIRVHSSNDSESESTFTSLVSPSF 1103
QY 1181 SGVTTGAVASANDOSHSTSVTNSNSVNTPTTLTSSQVTSSTSTNTFFIATVDSG 1240
DB 1104 SGIS--VYSTISPRHVNSTVTHLPSLSKPDVDPIDSSDAVTSTNDNSLTSLTSENGKT 1161
QY 1241 SIQHSWT 1248
DB 1162 SVAISTTF 1169

RESULT 5
US-60-096-409-16243
; Sequence 16243, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096, 409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 16243
; LENGTH: 1191
; TYPE: PRF
; ORGANISM: Candida albicans
US-60-096-409-16243

Query Match      64.5%; Score 4190.5; DB 27; Length 1191;
Best Local Similarity 68.1%; Pred. No. 1.6e-296;
Matches 863; Conservative 89; Mismatches 195; Indels 121; Gaps 19;

QY 1 MLOQFTLLFLYLISIAKTTTGVDSPNSLTWSNAANYAFKPGYPTWNAVLGWSLDGTS 60
DB 3 MLOQFTLLFLYLISIAKTTTGVDSPNSLTWSNAANYAFKPGYPTWNAVLGWSLDGTS 62
QY 61 ANPGDTFTLNMPCVFYKTTTSQTSVDLTADGKVATCFYSGEEFTTSTLCTVNDALKS 120
DB 63 ANPGDTFTLNMPCVFYKTTTSQTSVDLTADGKVATCFYSGEEFTTSTLCTVNDALKS 122
QY 121 SIKAFGTVTLPIAFNVGGTSSDLEDSKCFCTAGTNTVTENDGDKDISIDVEFEKSTVDP 180
DB 123 SIKAFGTVTLPIAFNVGGTSSDLEDSKCFCTAGTNTVTENDGDKDISIDVEFEKSTVDP 182
QY 181 SAYLYASRVMPSLNKKVTTLEVAPOCENGYSGTMGFSSNGDVAIDCSNIHIGITKGLND 240

```


Db 183 SGLYASRYMPLSLNKTTLFVAPQCENGYSCTMGFSSNGGVAIDCSNIHIGITKGLND 242
QY 241 WNPVSSSEFSYKTKCTSGIOIKYONVPAGYRPFIDAVISATDNQYTLAYTNDVTCAG 300
Db 243 WNPVSSSEFSYKTKCTSGIOIKYONVPAGYRPFIDAVISATDNQYTLAYTNDVTCAG 302
QY 301 SRLQSPFTLRWGTGKNSDAGSNGIVIVATRTVTDSTTAVATLTPNPVSDKTKTIELIQ 360
Db 303 SRLQSPFTLRWGTGKNSDAGSNGIVIVATRTVTDSTTAVATLTPNPVSDKTKTIELIQ 362
QY 361 PIPTTTITTSYGVVTSYLTAKTAPGETATVIVDVPYHTTTVTTSWTGTTTTTRTNP 420
Db 363 PIPTTTITTSYGVVTSYLTAKTAPGETATVIVDVPYHTTTVTTSWTGTTTTTRTNP 422
QY 421 TSDIDVVQVPLPNTVSTTEYWSQSFATTTTAVAPPGTDTVLIIRPPNHTVTTTEYX 480
Db 423 TDYIDVVQVPLPNTVSTTEYWSQSFATTTTAVAPPGTDSVLIIRPPNHTVTTTEYX 482
QY 481 SQSFATTTTAVAPPGTDSVLIIRPPNHTVTTTEYWSQSFATTTTAVAPPGTDSVLIIR 540
Db 483 SQSATSSTVAPPGTDSVLIIRPPNHTVTTTEYWSQSFATTTTAVAPPGTDSVLIIR 542
QY 541 PNPVTTTEYWSQSFATTTTAVAPPGTDSVLIIRPPNHTVTTTEYWSQSFATTTTAV 600
Db 543 PNPVTTTEYWSQSFATTTTAVAPPGTDSVLIIRPPNHTVTTTEYWSQSFATTTTAV 602
QY 601 PPGGTDVLIIRPPNHTVTTTEYWSQSFATTTTAVPGSGTDTVLIIRPPNHTVTTTEYX 660
Db 603 PPGGTDVLIIRPPNHTVTTTEYWSQSFATTTTAVPGSGTDTVLIIRPPNHTVTTTEYX 662
QY 661 SOSYATTTTITAPPGTDSVLIIRPPNHTVTTTEYWSQSFATTTTAVPGGTDVLIIR 720
Db 663 SOSYATTTTITAPPGTDSVLIIRPPNHTVTTTEYWSQSFATTTTAVPGGTDVLIIR 722
QY 721 PNPVTTTEYWSQSFATTTTAVAPPGTDTVLIIRPPNHTVTTTEYWSQSFATTTTAV 780
Db 723 PNPVTTTEYWSQSFATTTTAVPGGTDVLIIRPPNHTVTTTEYWSQSFATTTTAV 782
QY 781 PPGGTDVLIIRPPNHTVTTTEYWSQSFATTTTAVPGSGTDTVLIIRPPNHTVTTTEYX 840
Db 783 PPGGTDVLIIRPPNHTVTTTEYWSQSFATTTTAVPGSGTDTVLIIRPPNHTVTTTEYX 816
QY 841 SS--DGMLLSSITLVE---SETTTELICSDGKCSRSSSGSIVTNPDSNESSIVTST 894
Db 817 TGPPG--TDVLIIRPPNHTVTTTEYWSQSFATTTTAVPGGTDVLIIRPPNHTVTTTEYX 853
QY 895 VPTASTMSDLSSTGDISATSDNVSKGVSVTETSVTIOTTPNPLSSSVTSLQLSS 954
Db 854 APPGCTATVLIIRPPNHTVTTTEYWSQSFATTTTAVPGGTDVLIIRPPNHTVTTTEYX 904
QY 955 IPSVSESESKVFTSGNDQSGTHDSQSTTEIEIVTTSTKVLPPVSSNTDLTSEPTN 1014
Db 905 SFTVTTE-----YWSQSFATTTTAVPGGTDVLIIRPPNHTVTTTEYX 941
QY 1015 TREQPTLLST--SNSITEDITTQPTGNDGNTSSTNPVATSTLASASEEDNKSGS 1072
Db 942 ---PVTTEYWSQSFATTTTAVPGGTDVLIIRPPNHTVTTTEYX 989
QY 1073 HESASTSLKPSMGNSGLTST----EIEATTTSPTEAPSVAGSCTDVTTEP--TDT-- 1124
Db 990 YATTTTAVAPPG-----TATVLIIRPPNHTVTTTEYWSQSFATTTTAVPGGTDVLI 1043
QY 1125 -REOPTLLSTTKTSELVATTAQTNENGKSPSDTLTSSLTGTGTSASTSANSELYT--- 1180
Db 1044 IREPPNHTVTTTEYWSQSFATTTTAVPGGTDVLIIRPPNHTVTTTEYX 1103
QY 1181 SSVTGCAVASANDQSHSTSVTNSNISVNTPOTTLSQVTSSTSTFTASTVDGSG 1240
Db 1104 SCSIS--VVTISRPHVYVNSTVTHLPSSLKSPVDIPSSDAVSTNDNSLTSLTSGENKGT 1161
QY 1241 STIQHSTW 1248
I: I:

Db 1162 SVAISTTF 1169
RESULT 6
US-10-179-131-7791
; Sequence 7791, Application US/10179131
; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLLING, JORK
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: 2976-4031
; CURRENT APPLICATION NUMBER: US/10/179,131
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 7791
; LENGTH: 828
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-179-131-7791
Query Match 63.2%; Score 4105; DB 25; Length 828;
Best Local Similarity 97.1%; Pred. No. 1.7e-290;
Matches 803; Conservative 12; Mismatches 12; Indels 0; Gaps 0;
QY 434 PNPVTTTEYWSQSFATTTTAVAPPGTDTVLIIRPPNHTVTTTEYWSQSFATTTTAV 493
Db 2 PNPVTTTEYWSQSFATTTTAVAPPGTDTVLIIRPPNHTVTTTEYWSQSFATTTTAV 61
QY 494 PGGTDVLIIRPPNHTVTTTEYWSQSFATTTTAVPGGTDVLIIRPPNHTVTTTEYX 553
Db 62 PGGTDVLIIRPPNHTVTTTEYWSQSFATTTTAVPGGTDVLIIRPPNHTVTTTEYX 121
QY 554 QSVATTTTAVAPPGTDSVLIIRPPNHTVTTTEYWSQSFATTTTAVAPGCTDVIIR 613
Db 122 QSVATTTTAVAPPGTDSVLIIRPPNHTVTTTEYWSQSFATTTTAVAPGCTDVIIR 181
QY 614 PNPVTTTEYWSQSFATTTTAVPGSGTDTVLIIRPPNHTVTTTEYWSQSFATTTTAV 673
Db 182 PNPVTTTEYWSQSFATTTTAVAPPGTDTVLIIRPPNHTVTTTEYWSQSFATTTTAV 241
QY 674 PGGTDVLIIRPPNHTVTTTEYWSQSFATTTTAVAPGCTDVIIRPPNHTVTTTEYX 733
Db 242 PGGTDVLIIRPPNHTVTTTEYWSQSFATTTTAVAPGCTDVIIRPPNHTVTTTEYX 301
QY 734 QSVATTTTAVAPPGTDSVLIIRPPNHTVTTTEYWSQSFATTTTAVAPGCTDVIIR 793
Db 302 QSVATTTTAVAPPGTDSVLIIRPPNHTVTTTEYWSQSFATTTTAVAPGCTDVIIR 361
QY 794 MSSKISTSSNDITSIIPSRPHVYVNSTSDLSSTFESSMNTPTSISSDGMLLSSITLV 853
Db 362 MSSKISTSSNDITSIIPSRPHVYVNSTSDLSSTFESSMNTPTSISSDGMLLSSITLV 421
QY 854 TSETTTTELICSDGKCSRSSSGIVTNPDSNESSIVTSTVPTASTMSDLSSTDGISA 913
Db 422 TSETTTTELICSDGKCSRSSSGIVTNPDSNESSIVTSTVPTASTMSDLSSTDGISA 481
QY 914 TSSDNVSKGVSVTTETSVTITQTTNPPLSSSVTSLTQLSSIPSVSESESKVFTTNGDN 973
Db 482 TSSDNVSKGVSVTTETSVTITQTTNPPLSSSVTSLTQLSSIPSVSESESKVFTTNGDN 541
QY 974 QSGTHDSQSTSTETIEIVTTSSKVLPPVYVNSTDLTSEPTNTREQPTTSLTSSNSTEDI 1033
Db 542 QSGTHDSQSTSTETIEIVTTSSKVLPPVYVNSTDLTSEPTNTREQPTTSLTSSNSTEDI 601
QY 1034 TTSQPTGNDGNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGNSGLTTS 1093
Db 602 TTSQPTGNDGNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGNSGLTTS 661

QY 1094 TEIEATTSPTTEAPSAVSSGTDVTEPTDTRQPTTLSTTKTNSLSELVATTQATNENG 1153
 DB 662 TEIEATTSPTTEAPSAVSSGTDVTEPTDTRQPTTLSTTKTNSLSELVATTQATNENG 721
 QY 1154 KSPSDDTLSSLTGTSASTSANSSELVTSQSVTGGAVASANDQSHSTSVTNSIVSNTP 1213
 DB 722 KSPSDDTLSSLTGTSASTSANSSELVTSQSVTGGAVASANDQSHSTSVTNSIVSNTP 781
 QY 1214 QTTLSQOVTSSSPSTNFTASTYDGSIIQHSITWLYGLITLISLFI 1260
 DB 782 QTTLSQOVTSSSPSTNFTASTYDGSIIQHSITWLYGLITLISLFI 828

RESULT 7
 US-60-385-568-218
 ; Sequence 218, Application US/60385568
 ; GENERAL INFORMATION:
 ; APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Glandong Zeng
 ; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their Use
 ; FILE REFERENCE: 032796-134
 ; CURRENT APPLICATION NUMBER: US/60/385,568
 ; PRIOR FILING DATE: 2002-06-05
 ; NUMBER OF SEQ ID NOS: 418
 ; SEQ ID NO 218
 ; TYPE: PRT
 ; ORGANISM: Candida albicans
 US-60-385-568-218

Query Match 62.8%; Score 4078.5; DB 27; Length 1047;
 Best Local Similarity 65.0%; Pred. No. 2.1e-288;
 Matches 822; Conservative 100; Mismatches 120; Indels 223; Gaps 18;

QY 1 MQQFTLLFLYLISIAAKTITGVDFSNLSLWSNAANAFKPGPTWNAVLGWSLDGTS 60
 DB 1 MQQFTLLFLYLISIAAKTITGVDFSNLSLWSNAANAFKPGPTWNAVLGWSLDGTS 60
 QY 61 ANPGDFTLLMPCVFKYTTQSVDLTADGVAYATCFYSGEEFTFSLTCTVNDALKS 120
 DB 61 ANPGDFTLLMPCVFKYTTQSVDLTADGVAYATCFYSGEEFTFSLTCTVNDALKS 120
 QY 121 SIKAFGTVPLAFNVGGTSGDLSKCFITAGTNTVTFNDGDKDISIDVFEKSTVDP 180
 DB 121 SIKAFGTVPLAFNVGGTSGDLSKCFITAGTNTVTFNDGDKDISIDVFEKSTVDP 180
 QY 181 SAYLASRVNPSLKNKVTLLFVAPQENGYSCTMGFSSNGDVAIDCSNIHIGITKGLND 240
 DB 181 KGLTDSRVIPSLNKVSTLFAVAPQANGYTSCTMGFANFYGDVQIDCSNIHIGITKGLND 240
 QY 241 WNPVSSSEFSYTKTCTSGNGIOIKYONVPAGYRPFIDAVISATDVOYTLATVNDYTCAG 300
 DB 241 WNPVSSSEFSYTKTCTSGNGIOIKYONVPAGYRPFIDAVISATDVOYTLATVNDYTCAG 300
 QY 301 SRLQSKPFLRWTKYKNSDAGSNGIVIVATTFTVDSITAVTTLFPNPSVDKTKTIEIQ 360
 DB 301 GYWRAPFTLRWTKYKNSDAGSNGIVIVATTFTVDSITAVTTLFPNPSVDKTKTIEIQ 360
 QY 361 PIPTTTITTSYGVVTSYLTAPIGETATVIVDPVPHYHTTTVTSEWGTCTITTTTRNP 420
 DB 361 PIPTTTITTSYGVVTSYLTAPIGETATVIVDPVPHYHTTTVTSEWGTCTITTTTRNP 420
 QY 421 TSDIDTVVQVPLNPTVTSTYWSQSFATTTTVPAGGTDVLIIRPPNHTVTTTTEYW 480
 DB 421 TSDIDTVVQVPLNPTVTSTYWSQSFATTTTVPAGGTDVLIIRPPNHTVTTTTEYW 480
 QY 481 SOSFATTTTVPAGGTDVLIIRPPNHTVTTTTEYWSQSFATTTTVPAGGTDVLIIR 540
 DB 481 SESYTTTSTFTAPPGTDSVLIKEPPNPTVTTTEYWSQSFATTTTFTAPPGTDSVLI 540
 QY 541 PNPPTVTTEYWSQSFATTTTVPAGGTDVLIIRPPNHTVTTTTEYWSQSFATTTTVA 600

DB 541 PNPPTVTTEYWSQSFATTTTVPAGGTDVLIIRPPNHTVTTTTEYWSQSFATTTTVA 600
 QY 601 PPGGTDVLIIRPPNHTVTTTTEYWSQSFATTTTVPAGGTDVLIIRPPNHTVTTTTEYW 660
 DB 601 PPGGTDVLIIRPPNHTVTTTTEYWSQSFATTTTVPAGGTDVLIIRPPNHTVTTTTEYW 624
 QY 661 SOSYATTTTITAPPGETDVLIREPPNHTVTTTTEYWSQSFATTTTVPAGGTDVLIIR 720
 DB 625 SOSYATTTTITAPPGETDVLIREPPNHTVTTTTEYWSQSFATTTTVPAGGTDVLIIR 684
 QY 721 PNPHTVTTTEYWSQSFATTTTVPAGGTDVLIIRPPNHTVTTTTEYWSQSFATTTTVA 780
 DB 685 PNPHTVTTTEYWSQSFATTTTVPAGGTDVLIIRPPNHTVTTTTEYWSQSFATTTTVA 744
 QY 781 PPGGTDVLIIRPPNHTVTTTTEYWSQSFATTTTVPAGGTDVLIIRPPNHTVTTTTEYW 840
 DB 745 PPGGTDVLIIRPPNHTVTTTTEYWSQSFATTTTVPAGGTDVLIIRPPNHTVTTTTEYW 776
 QY 841 SSDGMLLSSTLVLTESETTELLICSDGKCSRLSSSGIVITNPDSNESSIVTSTVPTAST 900
 DB 777 -----LWSTWVETKTITETSCGDKGCSWVSSTRIVTIPNNIETPMVNTVDSITT 830
 QY 901 MSDLSSTDGISATSSDNVSKSGSVTTTETS-VTTIQTPNPLSSSVTSLTQLSSIPSVS 959
 DB 831 ESTS-QSPSGI-----FSESGSVETESTVTTAQTN-----PSVP 865
 QY 960 ESESKVTFTSNGDQSGHDSQSTSTEIEIVTTSKVLPPVVSNTDLTSEPTNTREQP 1019
 DB 866 TTESEVEFTTKNGNGNGPYESPSTH-----VKSMDENSEFT----- 902
 QY 1020 TTLSTTSNITEDITTSQPTGNDGNTSSTNPVPTVATSTLASAEEDNKGSHESASTS 1079
 DB 903 -----TSTAAS----- 908
 QY 1080 LKPSMGENSEGLTSTEIE---ATTTSPTEAPSAVSSGTDVTEPTDTRQPTTLSTTSK 1136
 DB 909 -----TSTDIEATATGSGVEASSPISSADET-----TTITTTAE 946
 QY 1137 TNSLVLATQATNEN-GKSPSTDLTSLTTGTSTASTANSELVTSQSVTGGAVASAND 1195
 DB 947 STSVI---EQPTNNGGKAPSA-----TSSPSTTTTANNDSVITG-----TTSN 989
 QY 1196 QSHSTSVTNSIVSNTPQTTLSQVTSSTPTNTFIATYDGSIIQHSITWLYGLITL 1255
 DB 990 QSQSQSQSN-----SDTQTTLSQVTSSTVSLVH--MLTFDGSQSVIQTWLCGLITL 1042
 QY 1256 LSLFI 1260
 DB 1043 LSLFI 1047

RESULT 8
 US-60-385-568-295
 ; Sequence 295, Application US/60385568
 ; GENERAL INFORMATION:
 ; APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Glandong Zeng
 ; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their Use
 ; FILE REFERENCE: 032796-134
 ; CURRENT APPLICATION NUMBER: US/60/385,568
 ; PRIOR FILING DATE: 2002-06-05
 ; NUMBER OF SEQ ID NOS: 418
 ; SEQ ID NO 295
 ; LENGTH: 1586
 ; TYPE: PRT
 ; ORGANISM: Candida albicans
 US-60-385-568-295

Query Match 60.8%; Score 3951; DB 27; Length 1586;
 Best Local Similarity 53.4%; Pred. No. 8.3e-279;
 Matches 859; Conservative 140; Mismatches 238; Indels 372; Gaps 34;

QY 1 MLQQTLLFLYLISIAAKTITGVDFSNLSLWSNAANAFKPGPTWNAVLGWSLDGTS 60

```

Db      1  MLLQFLLSLCVATAKVTIGVNSFDSLWTRAGNYAYKGNRPRTWNAVLGWSLDGTS 60
QY      61  ANPGDFTLNMPCVKRYTTSOTSDVLADGVKATCOFYSGEEFTTSTLTCTVNDALKS 120
Db      61  ANPGDFTLNMPCVKRYTTSOTSDVLADGVKATCOFYSGEEFTTSTLTCTVNDALKS 120
QY      121  SIKAGTDTLPIAFNNGVGTSSDLEDSKCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db      121  SIKALGTDTLPIAFNNGVGTSSDLEDSKCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
QY      181  SAYLZASRWPSLNVKTLFLVAPQCEGYTSGTMGFSNGDVAIDCSNIHIGITKGLND 240
Db      181  SGYFIASRLIPINKASITVAPQACANGYTSAGMGFTIGSGDITDCSNVHVGIKGLND 240
QY      241  WNPVSSSESFTTKTCTSGIOIKYONVPAGYRPFIDAVISATDNQYTLATNDYTCAG 300
Db      241  WNPVSSSLSLNKCTSGISITENVPAGYRPFIDAVISATDNQYTLATNDYTCAG 300
QY      301  SRLQSKPFLRWTKYKNSDAGSNGIVIVATRTVTDSTTAVTTLNPNPVDKTKTIEILO 360
Db      300  SLSQSKPFLRLRGYNSEANSNGFVIVATRTVTDSTTAVTTLNPNPVDKTKTIEILO 360
QY      361  PIPTTTITTSYGVVTSYTKTAPIGETATVIVDVPYHTTNTVTSWTGTTTTTRTNP 420
Db      360  PIPTTTITTSYGVVTSYTKTAPIGETATVIVDVPYHTTNTVTSWTGTTTTTRTNP 420
QY      421  TOSIDVNVQVLPNPTVSTTEYWSQSFATTTVTAPPGGDTVVIIRPPNHTVTTTEY 480
Db      420  TOSIDVNVQVLPNPTVSTTEYWSQSFATTTVTAPPGGDTVVIIRPPNHTVTTTEY 480
QY      481  SOSFATTTVTAPPGGDTVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGDTVVIIR 540
Db      480  SOSYATTTVTAPPGGDTVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGDTVVIIR 540
QY      541  PNPVTTTEYWSQSFATTTVTAPPGGDTVVIIRPPNHTVTTTEYWSQSFATTTVT 600
Db      540  PNPHTVTTTEYWSQSFATTTVTAPPGGDTVVIIRPPNHTVTTTEYWSQSFATTT 600
QY      601  PPGGDTVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGDTVVIIRPPNHTVTT 660
Db      600  PPGGDTVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGDTVVIIRPPNHTVTT 660
QY      661  SOSYATTTVTAPPGGDTVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGDTVVI 720
Db      660  SOSYATTTVTAPPGGDTVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGDTVVI 720
QY      721  PNPHTVTTTEYWSQSFATTTVTAPPGGDTVVIIRPPNHTVTTTEYWSQSFATTT 780
Db      720  PPNYTTVTTTEYWSQSFATTTVTAPPGGDTVVIIRPPNHTVTTTEYWSQSFATTT 780
QY      781  PPGGDTVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGDTVVIIRPPNHTVTT 840
Db      780  PPGGDTVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGDTVVIIRPPNHTVTT 840
QY      825  -----DLSTFE--SSSMNTPTSTIS-----SDGMLL-- 847
Db      840  SOSYATTTVTAPPGGDTVVIIRPPNHTVTTTEYWSQSFATTTVTAPPGGDTVVI 899
QY      848  -----SSTLVY-----ESETTTELICSDCKECSRLSS 875
Db      900  PNPVTTVTTTEYWSQSFATTTVTAPPGGDTVVIIRPPNHTVTTTEYWSQSFATTT 952
QY      876  SSGIVNPNDSNESSIV-----TSTVPTASTMSDLSLSTDCISAT----- 914
Db      953  TTTVTAPPGGDTVVIIRPPNHTVTTTEYWSQSFATTTVTATPGGDTVVIIRPPN 1012
QY      915  -----SSDNV-----SKGVSVTET----- 930
Db      1013  VTTTEYWSQSFATTTVTAPPGGDTVVIIRPPNHTVTTTEYWSQSFATTTVTAP 1072
QY      931  -----SVTT-----IQTPNPL-----SSSVT 947

```

RESULT 9

```

US-10-179-131-8318
; General Information:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLLING, JORK
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
; FILE REFERENCE: 2976-4031
; CURRENT APPLICATION NUMBER: US/10/179,131
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 8318
; LENGTH: 885
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-179-131-8318

Query Match      58.2%; Score 3777.5; DB 25; Length 885;
Best Local Similarity 79.0%; Pred. No. 1.6e-266;
Matches 707; Conservative 56; Mismatches 49; Indels 83; Gaps 3;

QY      1  MLOQFTLLFLYLSIASAKTITGVDFDSFNSLTWNAAYAPKPGCYPTWNAVLGWSLDGTS 60
Db      1  MLOQYTLFLYLSVATAKITGVFNSLTWNAATYKKGPTWNAVLGWSLDGTS 60
QY      61  ANPGDFTLNMPCVKRYTTSQTSVDLTADGVKATCOFYSGEEFTTSTLTCTVNDALKS 120
Db      61  ASPGDTFTLNMPCVKRYTTSQTSVDLTADGVKATCOFQAGEEFTTSTLTCTVNTLTP 120
QY      121  SKAFGTVTLPIAFNNGVGTSSDLEDSKCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db      121  SIKALGTVTLPIAFNNGVGTSSDLEDSKCTAGTNTVTFNDGDKKISINVDFERSNVDP 180

```

```

QY 181 SAYLYASRVMPSLNKVYTLFVAPQCENGYSGTMGFSSNGDAIDCSNIHIGITKGLND 240
D 181 KGYLTDTSRVPLSNKYSTLFVAPQCENGYSGTMGFANTYGVQIDCSNIHVGITKGLND 240
QY 241 WNPVSESESYTKTCSNGIQIKYONVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
D 241 WNPVSESESYTKTCSNGIFITYKNVPAGYRPFVDAISATDVNSYTLAYANEVTCAG 300
QY 301 SRLOSPFTLRWTKYKNSDAGSNGIVATRVTDSTTAVTTLPNPSVDKTKTIELQ 360
D 301 GYWRAPFTLRWTKYKNSDAGSNGIVATRVTDSTTAVTTLPNPSVDKTKTIELK 360
QY 361 PIPTTTITTSYGVGTTSYLTKTAPIGETATVIVDVPHYHTTTVTSEWGTITTTTTRNP 420
D 361 PIPTTTITTSYGVGTTSYLTKTAPIGETATVIVDVPHYHTTTVTSEWGTITTTTTRNP 420
QY 421 TDSIDTVVQVPLPNPTVSTTEYWSQSFATTTTAPPGGTDVLIIRPPNHNVTTEY 480
D 421 TDSIDTVVQVPLPNPTVSTTEYWSQSFATTTTAPPGGTDVLIIRPPNHNVTTEY 480
QY 481 SQSFATTTTVPAPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 527
D 481 SESITTTSTAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 527
QY 528 -----APPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 564
D 528 -----APPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 564
QY 565 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTTAPPGGTDVLIIRPPNHNVTTEY 624
D 565 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTTAPPGGTDVLIIRPPNHNVTTEY 624
QY 601 PPGETDVLIRPPNPTVTTEYWSQSFATTTTITAPPGGTDVLIIRPPNHNVTTEY 720
D 601 PPGETDVLIRPPNPTVTTEYWSQSFATTTTITAPPGGTDVLIIRPPNHNVTTEY 720
QY 625 SQSFATTTTVPAPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 779
D 625 SQSFATTTTVPAPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 779
QY 781 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTTAPPGGTDVLIIRPPNPTVTTEY 840
D 781 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTTAPPGGTDVLIIRPPNPTVTTEY 840
QY 780 -----APPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 823
D 780 -----APPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 823
QY 841 SQPYTTTTVIAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT-----SFSRPHYTNHTT 884
D 841 SQPYTTTTVIAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT-----SFSRPHYTNHTT 884

RESULT 10
US-60-385-568-297
; Sequence 297, Application US/60385568
; GENERAL INFORMATION:
; APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Giandong Zeng
; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their Use
; FILE REFERENCE: 032796-134
; CURRENT APPLICATION NUMBER: US/60/385,568
; CURRENT FILING DATE: 2002-06-05
; PRIOR FILING DATE: June 5, 2002
; NUMBER OF SEQ ID NOS: 418
; SEQ ID NO 297
; LENGTH: 886
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-385-568-297

Query Match 58.2%; Score 3777.5; DB 27; Length 886;
Best Local Similarity 79.0%; Pred. No. 1.6e-266;
Matches 707; Conservative 56; Mismatches 49; Indels 83; Gaps 3;

QY 1 MLOQFTLLFLYLSIAKAITGVDFSNLSLTWSNAATYKPGCTPTNNAVLGSLDGTG 60
D 1 MLOQFTLLFLYLSIAKAITGVDFSNLSLTWSNAATYKPGCTPTNNAVLGSLDGTG 60
QY 61 ANPGDFTLNMPCVFKYTTTSQTSVDLTADGKVATCOFYSGREFTTSTLTCTVNDALKS 120
D 61 ANPGDFTLNMPCVFKYTTTSQTSVDLTADGKVATCOFYSGREFTTSTLTCTVNDALKS 120
QY 61 ASPGDFTLNMPCVFKYTTTSQTSVDLTADGKVATCOFYSGREFTTSTLTCTVNDALKS 120
D 61 ASPGDFTLNMPCVFKYTTTSQTSVDLTADGKVATCOFYSGREFTTSTLTCTVNDALKS 120
QY 121 SIKAFGTVTLPIAFNVGGTSSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
D 121 SIKAFGTVTLPIAFNVGGTSSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
QY 121 SIKALGTVTLPIAFNVGGTSSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
D 121 SIKALGTVTLPIAFNVGGTSSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
QY 181 SAYLYASRVMPSLNKVYTLFVAPQCENGYSGTMGFSSNGDAIDCSNIHIGITKGLND 240
D 181 KGYLTDTSRVPLSNKYSTLFVAPQCENGYSGTMGFANTYGVQIDCSNIHVGITKGLND 240
QY 241 WNPVSESESYTKTCSNGIQIKYONVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
D 241 WNPVSESESYTKTCSNGIFITYKNVPAGYRPFVDAISATDVNSYTLAYANEVTCAG 300
QY 301 SRLOSPFTLRWTKYKNSDAGSNGIVATRVTDSTTAVTTLPNPSVDKTKTIELQ 360
D 301 GYWRAPFTLRWTKYKNSDAGSNGIVATRVTDSTTAVTTLPNPSVDKTKTIELK 360
QY 361 PIPTTTITTSYGVGTTSYLTKTAPIGETATVIVDVPHYHTTTVTSEWGTITTTTTRNP 420
D 361 PIPTTTITTSYGVGTTSYLTKTAPIGETATVIVDVPHYHTTTVTSEWGTITTTTTRNP 420
QY 421 TDSIDTVVQVPLPNPTVSTTEYWSQSFATTTTAPPGGTDVLIIRPPNHNVTTEY 480
D 421 TDSIDTVVQVPLPNPTVSTTEYWSQSFATTTTAPPGGTDVLIIRPPNHNVTTEY 480
QY 481 SQSFATTTTVPAPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 527
D 481 SESITTTSTAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 527
QY 528 -----APPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 564
D 528 -----APPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 564
QY 565 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTTAPPGGTDVLIIRPPNHNVTTEY 624
D 565 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTTAPPGGTDVLIIRPPNHNVTTEY 624
QY 601 PPGETDVLIRPPNPTVTTEYWSQSFATTTTITAPPGGTDVLIIRPPNHNVTTEY 720
D 601 PPGETDVLIRPPNPTVTTEYWSQSFATTTTITAPPGGTDVLIIRPPNHNVTTEY 720
QY 625 SQSFATTTTVPAPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 779
D 625 SQSFATTTTVPAPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 779
QY 781 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTTAPPGGTDVLIIRPPNPTVTTEY 840
D 781 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTTAPPGGTDVLIIRPPNPTVTTEY 840
QY 780 -----APPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 823
D 780 -----APPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 823
QY 841 SQPYTTTTVIAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT-----SFSRPHYTNHTT 884
D 841 SQPYTTTTVIAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT-----SFSRPHYTNHTT 884

```

```

Db 1 MLOQFTLLFLYLSIAKAITGVDFSNLSLTWSNAATYKPGCTPTNNAVLGSLDGTG 60
QY 61 ANPGDFTLNMPCVFKYTTTSQTSVDLTADGKVATCOFYSGREFTTSTLTCTVNDALKS 120
D 61 ANPGDFTLNMPCVFKYTTTSQTSVDLTADGKVATCOFYSGREFTTSTLTCTVNDALKS 120
Db 61 ASPGDFTLNMPCVFKYTTTSQTSVDLTADGKVATCOFYSGREFTTSTLTCTVNDALKS 120
QY 121 SIKAFGTVTLPIAFNVGGTSSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
D 121 SIKAFGTVTLPIAFNVGGTSSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db 121 SIKALGTVTLPIAFNVGGTSSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
QY 181 SAYLYASRVMPSLNKVYTLFVAPQCENGYSGTMGFSSNGDAIDCSNIHIGITKGLND 240
D 181 KGYLTDTSRVPLSNKYSTLFVAPQCENGYSGTMGFANTYGVQIDCSNIHVGITKGLND 240
QY 241 WNPVSESESYTKTCSNGIQIKYONVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
D 241 WNPVSESESYTKTCSNGIFITYKNVPAGYRPFVDAISATDVNSYTLAYANEVTCAG 300
QY 301 SRLOSPFTLRWTKYKNSDAGSNGIVATRVTDSTTAVTTLPNPSVDKTKTIELQ 360
D 301 GYWRAPFTLRWTKYKNSDAGSNGIVATRVTDSTTAVTTLPNPSVDKTKTIELK 360
QY 361 PIPTTTITTSYGVGTTSYLTKTAPIGETATVIVDVPHYHTTTVTSEWGTITTTTTRNP 420
D 361 PIPTTTITTSYGVGTTSYLTKTAPIGETATVIVDVPHYHTTTVTSEWGTITTTTTRNP 420
QY 421 TDSIDTVVQVPLPNPTVSTTEYWSQSFATTTTAPPGGTDVLIIRPPNHNVTTEY 480
D 421 TDSIDTVVQVPLPNPTVSTTEYWSQSFATTTTAPPGGTDVLIIRPPNHNVTTEY 480
QY 481 SQSFATTTTVPAPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 527
D 481 SESITTTSTAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 527
QY 528 -----APPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 564
D 528 -----APPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 564
QY 541 PNHTVTTTEYWSQSFATTTTAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 600
D 541 PNHTVTTTEYWSQSFATTTTAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 600
QY 565 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTTAPPGGTDVLIIRPPNHNVTTEY 624
D 565 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTTAPPGGTDVLIIRPPNHNVTTEY 624
QY 601 PPGETDVLIRPPNPTVTTEYWSQSFATTTTITAPPGGTDVLIIRPPNHNVTTEY 660
D 601 PPGETDVLIRPPNPTVTTEYWSQSFATTTTITAPPGGTDVLIIRPPNHNVTTEY 660
QY 625 SQSFATTTTVPAPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 684
D 625 SQSFATTTTVPAPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 684
QY 661 SQSYTTTTVIAPPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 720
D 661 SQSYTTTTVIAPPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 720
QY 685 PNHTVTTTEYWSQSFATTTTAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 744
D 685 PNHTVTTTEYWSQSFATTTTAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 744
QY 721 PNHTVTTTEYWSQSFATTTTITAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 780
D 721 PNHTVTTTEYWSQSFATTTTITAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 780
QY 745 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 779
D 745 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTT----- 779
QY 781 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTTAPPGGTDVLIIRPPNPTVTTEY 840
D 781 PPGGTDSVLIIRPPNPTVTTEYWSQSFATTTTAPPGGTDVLIIRPPNPTVTTEY 840
QY 780 -----APPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 823
D 780 -----APPGGTDVLIIRPPNPTVTTEYWSQSFATTTT----- 823
QY 841 SQPYTTTTVIAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT-----SFSRPHYTNHTT 884
D 841 SQPYTTTTVIAPPGGTDVLIIRPPNPTVTTEYWSQSFATTTT-----SFSRPHYTNHTT 884

RESULT 11
US-60-385-568-322
; Sequence 322, Application US/60385568
; GENERAL INFORMATION:
; APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Giandong Zeng
; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their Use
; FILE REFERENCE: 032796-134
; CURRENT APPLICATION NUMBER: US/60/385,568
; CURRENT FILING DATE: 2002-06-05
; PRIOR FILING DATE: June 5, 2002
; NUMBER OF SEQ ID NOS: 418
; SEQ ID NO 322
; LENGTH: 1756
; TYPE: PRT

```

ORGANISM: Candida albicans
us-60-385-568-322

```
Query Match      58.0%; Score 3765.5; DB 27; Length 1756;
Best Local Similarity 58.4%; Pred. No. 3.5e-265;
Matches 789; Conservative 136; Mismatches 260; Indels 165; Gaps 29;

Qy 1 MLOQFTLLFLYLSTASAKTITGVDFSNLSLWNAANYAFKPGYPTWNAVGLWSLDGTS 60
Db 1 MLLQFLLLSLVCVATAKVITGVFNSFDSLWTRAGNYAYKGNPTWNAVGLWSLDGTS 60

Qy 61 ANPGDTFLNMPCKFKYTTTSQTSVDLTADGVKYATCOFYSGEEFTTSLTCTVNDALKS 120
Db 61 ANPGDTFLNMPCKFKYTTTSQTSVDLTADGVKYATCOFYSGEEFTTSLTCTVNDALKS 120

Qy 121 SIKAGVTPLPIAFNVGSGTSDLESKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db 121 SIKALGVTLPIFNVGSGTSDLESQCFKAGTNTVTFNDGDKKISIDVDFETNEDA 180

Qy 181 SAYIYASRVMPSLNKTTLFVAPOCENGYSCTMGFSSNGDVAIDCSNIHIGITKGLND 240
Db 181 SGYFIASRLIPSINKVSIYVAPQCANGYTSANGFIVLTGDTTIDCSNVHVGITKGLND 240

Qy 241 WNPVSSSEFSYTKTCTSNGLIOIKYONVAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db 241 WNFVSSDLSYNKTSSTGISITYENVPAGYRPFDFVTSVSGQNR-QLRYTNDYACVG 299

Qy 301 SRLQSKPFTLWTKYKNSDAGSNGIVIVATRTVTDSTAVTTLFPNPSVDKTKTIELQ 360
Db 300 SLSQSKPFLNRLGNSEANSNGFIVATRTVTDSTAVTTLFPNPSVDKTKTIELQ 359

Qy 361 PIPTTTITTSVGVVTSYLTAPIGETATVIVDPVHTTTVTSEWGTITTTTTRNP 420
Db 360 PIPTTTITTSVGVVTSYLTAPIGETATVIVDPVHTTTVTSEWGTITTTTTRNP 419

Qy 421 TDSIDTVVQVPLNPNTVSTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 480
Db 420 TDSIDTVVQVPSNPNTVSTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 479

Qy 481 SOSFATTTTVPAGGTDVVIIRPPNHTVSTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 540
Db 480 SOSFATTTTVPAGGTDVVIIRPPNHTVSTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 539

Qy 541 PNPNTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVSTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 600
Db 540 PNPNTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVSTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 599

Qy 601 PPGGTDVVIIRPPNHTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 660
Db 600 PPGGTDVVIIRPPNHTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 659

Qy 661 SOSFATTTTVPAGGTDVVIIRPPNHTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 720
Db 660 SOSFATTTTVPAGGTDVVIIRPPNHTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 719

Qy 721 PNPNTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 780
Db 720 PNPNTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 779

Qy 781 PPGGTDVVIIRPPNHTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 827
Db 780 PPGGTDVVIIRPPNHTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYW 817

Qy 828 TFESSMNTPTSISS--DGMLLSSTLVE---SETTELICSDGKCEKSLSSSGGLVT 881
Db 838 -YWSQSFATTTTVPAGG---TATVIIRPPNHTVTTTEY-----WSSQSFATTTTVPAGG 886

Qy 882 NPDNESSIV-----TSVPTASTMSDLSLSTDGISATSDSNVSKSGSVVTESTVTTIQ 936
Db 887 GPPGTDVVIIRPPNHTVTTTEYSQSFATTTTVPAGG-----TATVIIR 934

Qy 937 TTPNPL-----SSSVTSLTQLSIPSVSES-----ESKVTFTSNGDQSGTHDSQS 982
Db 937 TTPNPL-----SSSVTSLTQLSIPSVSES-----ESKVTFTSNGDQSGTHDSQS 982
```

```
Db 935 EPPNPNTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTE-----YWSQS 986

Qy 983 TSTEIVTTSSTKVLPPVSSNTDLTSEPTNTRQPTTLST---SNSITEDITTSOPT 1039
Db 987 YAT-----TTTVPAGGTDVVIIRPPN-----YVTTTEYSQSFATTTTVPAGG 1033

Qy 1040 GDNGDNTSSTNPVTATSTLASASEEDNKSGSHESASTSLKPSMG-----ENSLG 1090
Db 1034 GGTDTVIIRPPSPPTVTTT-----EWSQSFATTTTVPAGGTDVVIIRPPNHTV 1085

Qy 1091 TTS---TEIATTTSTP-----EAPSPAV-----SSGTDVTTTEPTDT- 1124
Db 1086 TTEYSQSFATTTTVPAGGTDVVIIRPPSPPTVTTTEYSQSFATTTTVPAGGTDV 1145

Qy 1125 ---REQPT-TLSSTKTNSELVATQATNENG-----KSPSTDLTSSLTCTTSAST 1172
Db 1146 TVIIRPPNHTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYSQSF 1202

Qy 1173 SANSSELVTS--GSVTGGAVASASN-----DQSHSTSVT-----NSNSIVSNTP-- 1213
Db 1203 FATTTVPAGGTDVVIIRPPNHTVTTTEYSQSFATTTTVPAGGTDVVIIRPPN 1262

Qy 1214 QTTLSQVTSSTSPSTNTFIATYDGSII 1243
Db 1263 PTVTTTEYSQSFATTTTVPAGGTDVVIIRPPNHTVTTTEYSQSFATTTTVPAGG 1292
```

RESULT 12

```
US-10-179-131-7998
; Sequence 7998, Application US/10179131
; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHINER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: 2976-4031
; CURRENT APPLICATION NUMBER: US/10/179,131
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 7998
; LENGTH: 1182
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-179-131-7998
```

```
Query Match      57.6%; Score 3739; DB 25; Length 1182;
Best Local Similarity 62.2%; Pred. No. 1.7e-263;
Matches 770; Conservative 119; Mismatches 227; Indels 122; Gaps 24;
```

```
Qy 1 MLOQFTLLFLYLSTASAKTITGVDFSNLSLWNAANYAFKPGYPTWNAVGLWSLDGTS 60
Db 9 MLLQFLLLSLVCVATAKVITGVFNSFDSLWTRAGNYAYKGNPTWNAVGLWSLDGTS 68

Qy 61 ANPGDTFLNMPCKFKYTTTSQTSVDLTADGVKYATCOFYSGEEFTTSLTCTVNDALKS 120
Db 69 ANPGDTFLNMPCKFKYTTTSQTSVDLTADGVKYATCOFYSGEEFTTSLTCTVNDALKS 128

Qy 121 SIKAGVTPLPIAFNVGSGTSDLESKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db 129 SIKALGVTLPIFNVGSGTSDLESKCFKAGTNTVTFNDGDKKISIDVDFETNEDA 188

Qy 181 SAYIYASRVMPSLNKTTLFVAPOCENGYSCTMGFSSNGDVAIDCSNIHIGITKGLND 240
Db 189 SGYFIASRLIPSINKVSIYVAPQCANGYTSANGFIVLTGDTTIDCSNVHVGITKGLND 248

Qy 241 WNPVSSSEFSYTKTCTSNGLIOIKYONVAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db 249 WNFVSSDLSYNKTSSTGISITYENVPAGYRPFDFVTSVSGQNR-QLRYTNDYACVG 307
```

QY 301 SRLQKPFLLRWTGYKNSDAGSNGIVIVATRTVIDSTAVTTLPNPSVDKTKTIEILQ 360
Db 308 SSIQSKPFLRLRGYNSEANSNGFIVATRTVIDSTAVTTLPNPSVDKTKTIEILQ 367
QY 361 PIPTTTTYSYGVVTSYLTAKPIGETATVIVDVPYHTTTVTSEWTKTITTTTRTP 420
Db 368 PIPTTTTYSYGVVTSYLTAKPIGETATVIVDVPYHTTTVTSEWTKTITTTTRTP 427
QY 421 TDSIDVNVQVPLPNTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEY 480
Db 428 TDSIDVNVQVPLPNTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEY 487
QY 481 SOSFATTTTAPPGTDSVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDSVVIIR 540
Db 488 SOSYATSSIVTAPPGTDSVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDSVVIIR 547
QY 541 PNPNTVTTEYWSQSFAATTTTAPPGTDSVVIIRPPNHTVTTEYWSQSFAATTTT 600
Db 548 PNPNTVTTEYWSQSFAATTTTAPPGTDSVVIIRPPNHTVTTEYWSQSFAATTTT 607
QY 601 PPGTDTVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEY 660
Db 608 PPGTDTVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEY 667
QY 661 SOSYATTTTITAPPGTDTVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDTVVIIR 720
Db 668 SOSYATTTTITAPPGTDTVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDTVVIIR 727
QY 721 PNPNTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEYWSQSFAATTTT 780
Db 728 PNPNTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEYWSQSFAATTTT 787
QY 781 PPGTDTVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEY 827
Db 788 PPGTDTVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEY 845
QY 828 TFESSMNTPTSISS--DGMLLSSTPLVTE---SETTELICSDGKESRLSSSSGIVT 881
Db 846 -YWSQSFAATTTTAPPGT--TATVVIIRPPNHTVTTEY--WSQSFAATTTT 894
QY 882 NPDSSNESSIV-----TSVTPTASTMDSLSSTGTSATSSDNVSKGVSVTTSVT 936
Db 895 GPPGGTDTVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEY 942
QY 937 TTPNPL-----SSSVTSLSLTSSIPSYSES-----ESKVTFTSNGDNGSHDSOS 982
Db 943 EPPNPTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEY--YWSOS 994
QY 983 TSTEIEIVTSTKVLPPVSSNTDLTSEPTNTRQOPTLSLT---SNSITEDITTSQPT 1039
Db 995 YAT-----TITVTAPPGTATVVIIRPPN-----YVVTTEYWSQSFAATTTT 1041
QY 1040 GONGDNSTSNPVTPTATSLASSEDNKSGBESASTSLKPSMGENSEGLTST-----E 1095
Db 1042 GGTDTVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEY 1087
QY 1096 IEATTTSPTEAPSVSSGTDVTEP--TDT---REQPT--TLSTSKTNSLVAITQATN 1149
Db 1088 PNPNTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEYWSQSFAATTTT 1147
QY 1150 ENGG-----KSPSTDLSLSTTGTSTASTSANSSELVT 1180
Db 1148 PPGTATVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEY 1182

RESULT 13
US-10-179-131-10126
; Sequence 10126, Application US/10179131
; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.

; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLLING, JORK
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
; FILE REFERENCE: 2976-4031
; CURRENT APPLICATION NUMBER: US/10/179,131
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 10126
; LENGTH: 1581
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-179-131-10126

Query Match 57.5%; Score 3736.5; DB 25; Length 1581;
Best Local Similarity 58.4%; Pred. No. 4e-263;
Matches 786; Conservative 128; Mismatches 264; Indels 167; Gaps 29;

QY 7 LFLFLYLSIASAKTITTCVDFDSFNSLWTSNAANAFKPGYTNNAVLGWSLDGTSANPGDT 66
Db 13 VHIIFSDVCKNFITGVFNSFDSLWTRAGNAYKGNPRTWNAVLGWSLDGTSANPGDT 72
QY 67 FTLNPCVFKYTTTSOTSDLTADGVKYATCQFYSGEFTTFTSLTCTVNDALKSSSKAFG 126
Db 73 FTLNPCVFKYTTSDTSDLTADGVKYATCQFYSGEFTTFTSLTCTVNDALKSSSKAFG 132
QY 127 TTVPLTAFNVGGTSGTDLSDKCFAGTNTVTFTNDGDKDISIDVFEKSTVDPSAYLYA 186
Db 133 TTVPLTAFNVGGTSGTDLSDKCFAGTNTVTFTNDGDKDISIDVFEKSTVDPSAYLYA 192
QY 187 SRVMSPLNKTFLVAPQCENGYTSCTMGFSNNGDVAIDCSNIHIGIKGLNDWNPVS 246
Db 193 SRLIPSINKVSTIYVAPQCENGYTSCTMGFSNNGDVAIDCSNIHIGIKGLNDWNPVS 252
QY 247 SESFSTKCTCTNGIQKYNQVAGYRPFIDAYISATDVNOYTLATNDYTCAGSLQSK 306
Db 253 SDLSYNKCTCTNGIQKYNQVAGYRPFIDAYISATDVNOYTLATNDYTCAGSLQSK 311
QY 307 PFTLRWTKYKNSDAGSNGIVIVATRTVIDSTAVTTLPNPSVDKTKTIEILQIPTTT 366
Db 312 PFNLRLRGYNSEANSNGFIVATRTVIDSTAVTTLPNPSVDKTKTIEILQIPTTT 371
QY 367 ITTSYGVVTSYLTAKPIGETATVIVDVPYHTTTVTSEWTKTITTTTRTPNPTSD 426
Db 372 ITTSYGVVTSYLTAKPIGETATVIVDVPYHTTTVTSEWTKTITTTTRTPNPTSD 431
QY 427 VVQVPLPNTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEYWSQSFAAT 486
Db 432 VVQVPLPNTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEYWSQSFAAT 491
QY 487 TTTVTAPPGTDSVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDSVVIIRPPNHT 546
Db 492 TTTVTAPPGTDSVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDSVVIIRPPNHT 551
QY 547 TTTTEYWSQSFAATTTTAPPGTDSVVIIRPPNHTVTTEYWSQSFAATTTTAPPG 606
Db 552 TTTTEYWSQSFAATTTTAPPGTDSVVIIRPPNHTVTTEYWSQSFAATTTTAPPG 611
QY 607 TTVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDSVVIIRPPNHTVTTEYWSQS 666
Db 612 SVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDSVVIIRPPNHTVTTEYWSQS 671
QY 667 TTTTAPPGTDTVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHT 726
Db 672 TTTTAPPGTDTVVIIRPPNHTVTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHT 731
QY 727 TTTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEYWSQSFAATTTTAP 786
Db 732 TTTTEYWSQSFAATTTTAPPGTDTVVIIRPPNHTVTTEYWSQSFAATTTTAP 791
QY 787 TWIIE-----SMSSSKISTSSNDITSIPFSRPHYVNSTTDLSTFESS 832

Db	792	TVIIREPPSPVTTTTEYWSQYATTTVTAPPGGTATVIRREPPNYTTTTE	YWSQ	847
Qy	833	SNMTPTSLS--DGLLSSTLVTB---SETTELICSDGKESRLSSSGIVTNPDN	886	
Db	848	SVATTTVTGPPGG--TDTVIIRPPNPVTITTEY-----WSQFATTTVTAPGG	897	
Qy	887	ESSIV-----TSTVPTASTMSDLSLSDGICSATSDNYSKGVSTTETSVTIQTTPNP	941	
Db	898	TDSVIIRPPNPVTITTEYWSQYATTTVTAPPGG-----TDSVIIRPPNP	945	
Qy	942	L-----SSVTSITQLSSIPSVSES-----ESKVTFTSNGDNQSDHDSOSTTEI	987	
Db	946	TVTTEYWSQYATTTVTAPPGGDSVIIRPPNPVTITTE-----YWSQYATTT	997	
Qy	988	ELVTSSTKVLPPVSSNNDLTSEPTNREOQPTTLSTT---SNSITEDITTQPTGDNGD	1044	
Db	998	TVTATPG-----GTDVIIRPPN-----YTVTTEYWSQYATTTVTAPPGGDS	1044	
Qy	1045	NTSSTNPVTATSLASASEEDNKSGSHESASTSLKPSMGNSGL-----TSTTEI	1096	
Db	1045	VIIRPPNPVTITTE-----EYWSQYATTTVTAPPGGDSGIIRPPNPSTVTTEY	1096	
Qy	1097	-----EATTSPT-----EAPSPAVS-----GTDVTTEP--TDT--R	1125	
Db	1097	WSQYATTTVTAPPGGDSVIIRPPDPTVTSTEYWSQFATTTVTAPPGGDSVIIR	1156	
Qy	1126	EQPT--TLSTKTSNSELVATTOATNENG-----KSPSTDLSLTTGTSASTSANSE	1177	
Db	1157	EPNPVTITTEYWSQYATTTVTAPPGGDSVIIRPPNPVT---TTEYWSQFATTT	1213	
Qy	1178	LVTS--GSVTGGAVASASN-----DQSHSTSVT-----NSNSIVSNTP--QTLS	1218	
Db	1214	TVTAPPGGDSVIIRPPNPVTITTEYWSQYATTTVTAPPGGDSVIIRPPNPVTIT	1273	
Qy	1219	QOVTSSSPSTNTFIASDTGSGSII	1243	
Db	1274	TEYWSQYATTTVTAPPGGTATVI	1298	
RESULT 14				
; Sequence 214, Application US/60385568				
; GENERAL INFORMATION: Willins, Arturo Morales, Guillaume Cottarel, Giandong Zeng				
; APPLICANT: Deborah Willins, Identification of Candida Cell Surface Proteins and their use				
; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and their use				
; FILE REFERENCE: 032796-134				
; CURRENT APPLICATION NUMBER: US/60/385,568				
; CURRENT FILING DATE: 2002-06-05				
; PRIOR FILING DATE: June 5, 2002				
; NUMBER OF SEQ ID NOS: 418				
; SEQ ID NO 214				
; LENGTH: 1523				
; TYPE: PRT				
; ORGANISM: Candida albicans				
US-60-385-568-214				
Query Match				
Best Local Similarity 57.2%; Score 3716.5; DB 27; Length 1523;				
Matches 798; Conservative 148; Mismatches 300; Indels 289; Gaps 30;				
Qy	1	MLQOFTLLFLYSIASATITGVDFDSFNLSLWSNAANYAFKPGYPTWNAVLGWSLDGTS	60	
Db	1	MLLQFLLSLVCVSATAKVIICGFDSFNLSLWNAASYRGPNATWTAIGWSLDGAT	60	
Qy	61	ANPGDTFLNMPCKVFKYTSQSDVLDTADGVKYATCFQYSGEETFTSTLCTVNDALKS	120	
Db	61	ASAGDTFLDMPCKVFKFTDQTSIDLVDAGRTYATCNLSAEETFTSSVSCVTITMTA	120	
Qy	121	SIKAGTFTLPIAFNVGSGTSDLEDSKCFAGTAGTNTVTFNDGDKDISIDVEFEKSTVDP	180	
Db	121	DTKAGTFTLPIFSVGGSGSDVLANSQCFTAGTNTVTFNDGDTSTISATVDFEKSSTVAS	180	


```

1255 VVTRTA' ATPTPIIGDNGSGKSGELSGTGVTTNTATPDVPSTKVPNPGAPGTG 1314
1086 EV" FEIEATTTSPTEADSPAVSSGTDVTTTEPTDT-----REOPTTL 1131
1315 PSTETQTNVPGSPNIPATGT-TDIIRESTTVSHVTGNGNTGVPMPNPALT 1373
11 KTNSELVATQATNE-----NGKSPSTDLS--SLTGTSASTSANSSELVTSGS 1183
1 SLTG-----ATNSATNPSETGVNTGSRGTNIIVTPPSSATATVVPCTDNGATKQG 1429
1184 VVGAVAS-----ASNDQSHSTSVTNSISIVNTQTLLSQQVTS 1224
1430 DTAGNSNGFTATNTQGNNEPQNPGNTGCEPVGTTGTOSVESTOPTLLSQQVTS 1489
1225 STSTNTFIATDVGSGSIHQHSTWLYGLITLLSLF 1259
1490 LISTP--LASTPDGSGSIQVHSGWLVLVLTATISIF 1522

RESULT 15
US-60-385-568-323
; Sequence 323, Application US/60385568
; GENERAL INFORMATION:
; APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Giandong Zeng
; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their
; FILE REFERENCE: 032796-134
; CURRENT APPLICATION NUMBER: US/60/385,568
; CURRENT FILING DATE: 2002-06-05
; PRIOR FILING DATE: June 5, 2002
; NUMBER OF SEQ ID NOS: 418
; SEQ ID NO 323
; LENGTH: 1593
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-385-568-323

```

Query Match	55.7%	Score 3620	DB 27	Length 1593
Best Local Similarity	53.9%	Pred. No. 1.3e-254		
Matches 763	Conservative 158	Mismatches 308	Indels 186	Gaps 25
QY	1	MLQOFTLLFLYLISASAKTITGVDFNSLTWNAANAAPKPGYPTWNAVLGWSLDGTS	60	
DB	1	MLLOFLLLLSCVSAKAVKVTIGIFDSFNSLTWTNAASYSYRGANPTWTAVIGWSLDGAT	60	
QY	61	ANPGDTFTLNMPCVFKYKTSQTSVDLTADGVKATQCFYSGBEFTFTSTTCTVNDALKS	120	
DB	61	ASAGDTFTLDMPCVFFKFTDQTSIDLVADGRTYATCNLSAAEFTTFSSVCIVTTMTA	120	
QY	121	SIKAFGTVTLPIAFNVTGGTSSDLEDSKCTAGTNTVTFNDGDKDISDVEFEKSVDP	180	
DB	121	DTKAIGTVTLPFSEFVGSGSDVDLANSQCFTAGINTVTFNDGDTSTSTVDFEKSIVAS	180	
QY	181	SAYLYASRWMPSLNKVTTLEVAPOCENGYSYTGMGFSSNGDVAIDCSNIHIGITKGLND	240	
DB	181	SDRILLRIPLSLSQAVSLFLPOECANGYTSGMTGFSTAGTGATIDCSNTHVHGISGLND	240	
QY	241	WNYPVSESFSYKTCFSNGIQIKYQNPAGYRPFIDAYISATDVNOYTLAYTNDYTCAG	300	
DB	241	WNPISESFSYKTCSTSTSVLYTIYQNPAGYRPFVDAYYSAFVSSYAMRYNIYACVG	300	
QY	301	SRLOSKRPTLWRVTGYKNSDAGSGIVATATRTVTDSTTAVTTLPPFNPSSVDKKTIELQ	360	
DB	301	AASVDDSFTHWGSYNSQAGSGNITIVATRTVTDSTTAVTTLPPFNPSSVDKKTIELQ	360	
QY	361	PIPTTTTITSYGVVTTSYLAKTPAGTATVVDVPYHTTTTIVTSEWGTITTTTTTRNP	420	
DB	361	PIPTTTTITSYGVVTTSTKATPAGTATVVDVPYHTTTTIVTSEWGTITTTTTTRNP	420	
QY	421	TDSIDTVVGVPLPNPTVSTTEYWSQSFATTTTTAPGGTDTVIIIREPPNHVTTTTTEYV	480	
DB	421	TDSIDTVVGVPLPNPTVSTTEYWSQSFATTTTTAPGGTDTVIIIREPPNHVTTTTTEYV	480	

Search completed: June 11, 2003, 17:20:09
Job time : 309 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:13:36 ; Search time 58 Seconds
(without alignments)
2242.806 Million cell updates/sec

Title: US-09-715-876-8

Perfect score: 6495

Sequence: 1 MLOQFTLLFLYLSIASAKTI.....SIHQSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PublishedApplications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	939	14.5	1537	10	US-09-801-368-104
2	932.5	14.4	1322	10	US-09-801-368-114
3	721.5	11.1	5179	9	US-10-025-380-1068
4	721.5	11.1	5179	10	US-09-801-368-1068
5	721.5	11.1	5179	10	US-09-801-368-1068
6	711	10.9	1075	10	US-09-801-368-110
7	675.5	10.4	1367	10	US-09-801-368-108
8	640.5	9.9	2344	10	US-09-815-242-12713
9	545.5	8.4	957	9	US-10-025-380-1065
10	545.5	8.4	957	10	US-09-801-368-1065
11	545.5	8.4	957	10	US-09-801-368-1065
12	525.5	8.0	5877	9	US-10-142-515-11
13	501	7.7	1169	10	US-09-801-368-106
14	463	7.1	688	10	US-09-864-761-36047
15	428	6.6	1236	9	US-09-769-787-109
16	417	6.4	560	10	US-09-815-242-13057
17	392	6.0	502	10	US-09-815-242-5904
18	390	6.0	596	9	US-10-063-547-100
19	390	6.0	596	9	US-10-174-590-310

20	390	6.0	596	9	US-10-176-758-310	Sequence 310, App
21	390	6.0	596	9	US-10-063-616-100	Sequence 100, App
22	390	6.0	596	9	US-10-175-737-310	Sequence 310, App
23	390	6.0	596	9	US-10-063-502-100	Sequence 100, App
24	390	6.0	596	9	US-10-173-708-310	Sequence 310, App
25	390	6.0	596	9	US-10-175-738-310	Sequence 310, App
26	390	6.0	596	9	US-10-175-752-310	Sequence 310, App
27	390	6.0	596	9	US-10-176-482-310	Sequence 310, App
28	390	6.0	596	9	US-10-176-757-310	Sequence 310, App
29	390	6.0	596	9	US-10-176-913-310	Sequence 310, App
30	390	6.0	596	9	US-10-180-552-310	Sequence 310, App
31	390	6.0	596	9	US-10-180-557-310	Sequence 310, App
32	390	6.0	596	9	US-10-173-700-310	Sequence 310, App
33	390	6.0	596	9	US-10-174-572-310	Sequence 310, App
34	390	6.0	596	9	US-10-174-579-310	Sequence 310, App
35	390	6.0	596	9	US-10-174-582-310	Sequence 310, App
36	390	6.0	596	9	US-10-174-588-310	Sequence 310, App
37	390	6.0	596	9	US-10-175-739-310	Sequence 310, App
38	390	6.0	596	9	US-10-175-740-310	Sequence 310, App
39	390	6.0	596	9	US-10-175-743-310	Sequence 310, App
40	390	6.0	596	9	US-10-176-488-310	Sequence 310, App
41	390	6.0	596	9	US-10-176-492-310	Sequence 310, App
42	390	6.0	596	9	US-10-176-747-310	Sequence 310, App
43	390	6.0	596	9	US-10-176-750-310	Sequence 310, App
44	390	6.0	596	9	US-10-176-985-310	Sequence 310, App
45	390	6.0	596	9	US-10-176-987-310	Sequence 310, App

ALIGNMENTS

RESULT 1
US-09-801-368-104
; Sequence 104, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 104
; LENGTH: 1537
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-104

Query Match 14.5%; Score 939; DB 10; Length 1537;
Best Local Similarity 25.8%; Pred. No. 1.1e-37;
Matches 414; Conservative 246; Mismatches 520; Indels 426; Gaps 84;
QY 2 LQOFTLLFLYLSIASAKT-----ITGVDFDSF-----NSLTWNAANYAF----- 40
DB 10 LAVFTLLAL-TSVASGATEACLPAGQRKSGMNIYQYSLKDSSTYSNAAYMAYGASKT 68

QY 41 -----KGGYPTW-----NAVIGWLD--G 58
Db 69 KLGSGVGGQTDISIDYNPCVSSSGFPQEDSYGNWCKGMGACNSOGIAYWSTDLFG 128
QY 59 TSANPGDFTLLM-----PCVFKTTQ-----TSVD 85
Db 129 FYTTPTNV-TLEMTGYFLPQGSYTFKFAFATVDDSAILSVGATAFNCAOQPPITSTN 187
QY 86 LTADGVK-----YATCOFYSGEBFTTSTLTCTVNDALKSSIKAFGT-----VTLF 131
Db 188 FTIDGKPMGWSLPPNIEGTVMYAGYPM-----KVYSNVSWGTLPISVLP 238
QY 132 IAFNVGGTSSDLEDSKCFAGTNTVFNDDGDKDISIDVEKSTV-DPSAYLYASRYM 190
Db 239 -----DGTVSDDEG-----VYISFDD-----DLSQSNCTVPDPSNYA-VSITT 277
QY 191 PSLNKVTLFLVAPOCENGYSCTMGFSNGDVAIDCSNIHIGIKLNDNYPVSSRSF 250
Db 278 TTTEPWGTFTSTSTEMTGTGNGVPTDETVIVIRTPPTASTIITTEPWNSTFTST 337
QY 251 SYTKCTSGNIGIQYQNPAGYRPFIDAYISATDVNOVTLAYTNDVTCAGSRLQSKPFTL 310
Db 338 ELTTVTGNGVTDII-----VIRPTTATTITTEPWNSTFTSTEL-----384
QY 311 RWTGKNSDAGSNGI---VIVATRTVDSSTFAVTL-PFNPSVDKTKT-----355
Db 385 -----TTVTGNGLPDDETVIVIRPTTATTATTQPNWDTFTSTELTTVTGTNGL 438
QY 356 -----IEIQIPITTT-----SVGVTTSYLTKTA-----PIGETATVIVDVPH 398
Db 439 PDETVIVIRPTTATTATTQPNWDTFTSTELTTVTGTNGLPDTDET-IIVIRPTT 497
QY 399 TTT-TVTSEWGTIT-----TTTRN-PTDSIDTVWQVPLPNTVST-----EY 443
Db 498 ATAMTTQPNWDTFTSTELTTVTGTNGLPDT-ETIIV---IRPTTATTATTQTP 552
QY 444 WSQSFAIT-----TTVTAPP-GTD--TVIIRPPNHT--VTTTEYWSQSFAIT-----TTV 490
Db 553 WNDTFTSTSTEMTGTGNGLPDDETVIVIRPTTATTITTEPWNSTFTSTEMTIV 612
QY 491 TAPP-GTDS--VIIRPPNPT--VTTTEYWSQSFAIT-----TTVTAPP-GTDS--VII 538
Db 613 TGTNGLPDTDETVIVIRPTTATTATTQPNWDTFTSTSTEMTGTGNGLPDDETVI 672
QY 539 REPPNPT--VTTTEYWSQSFAIT--TTVTAPP-GTDS--VIIRPPNHT--VTTTE 587
Db 673 RPTTATTATTQPNWDTFTSTELTTVTGTGLPDDETVIVIRPTTATTATTQTP 732
QY 588 WSQSFAIT-----TTVTAPP-GTD--TVIIRPPNHTV--TTEYWSQSFAIT-----TTV 634
Db 733 WNDTFTSTSTEMTGTGNGVPTDETVIVIRPTSEGLISTTTEPWTGFTSTSTEMTIV 792
QY 635 T--GPPSGTDVIIRPPNP--TVTTTEYWSQSFAIT--TTTAP--PGEDTVLI 682
Db 793 TGTNGOPTDETVIVIRPTSEGLISTTTEPWTGFTSTSTEMTGTGNGVPTDETVI 852
QY 683 REPPNHTV--TTEYWSQSFAIT--TTVTAPP-GTD--TVIIRPPNHTV--TTEY 731
Db 853 RPTSEGLISTTTEPWTGFTSTSTEMTGTGNGOPTDETVIVIRPTSEGLISTTTEP 912
QY 732 WSQSFAITTTVTAPP-GTD-----TVIIRPPNHTV--TTEYWSQSFAITTTVTAPP 782
Db 913 WGTGFTSTSTEMTGTGNGVPTDETVIVIRPTSEGLISTTTEPWTGFTSTSTEMTIV 972
QY 783 GGT-----DTVIIYESMSSKI-----STSSNDITSIIPEFSRP-----816
Db 973 TGTNGOPTDETVIVIRPTSEGLISTTTEPWTGFTSTSTEMTGTGNGOPTDETVI 1032
QY 817 -----HYVNSTSD--LSTFSSSMNTPTSISDGMLLSSSTPLVPESETT-----FELICS 865
Db 1033 RPTSEGLVTTTTEPWTGFTSTSTEMTGTGNGLPDDETVIVIRPTTATLSSLSSSS 1092
QY 866 DGKESRLSSSGIVTNP--DSNESSIVTSTVPTASTMSDLSLSDGTSATSSDNVSKSG 923

Db 1093 SQOITSSITSSRIIT-PFIPNGTSVISSSVISSVTSSLTSSPVISSSVISSSTTTS 1151
QY 924 VSVTTETSVTIIQTTPNPLSSSVTSLTQLSSPSVSESKVFTFTNGDNQSGTHDSQT 983
Db 1152 TSIFSESSKSV--IPTSSSTSGSESETSSAGSVSS--SFISSESSKSPYSSSS- 1204
QY 984 SPEIIVITSSP-----KVLPPVSSNTDLTSEPTNTREOPTTLSTTS-----NSITE 1031
Db 1205 ---LPLVTSATISQETASSLPATTT-----TKTSEQTLTVVTSCSHVCTESTSP 1252
QY 1032 DITTSOPTCDNGDNTSSN--PVPTVATSTLASAEEDNKSGSHESASTSLKPSMGNSG 1089
Db 1253 AIVSTATVSGVTEYTTWCPISTTETTKQTKTEQTETTKQTVVTVISSCESDVCS 1312
QY 1090 LTTSTEIEATITSPAPSPAVSSGTDVTEPTDREQPTTLSTT-----SKTNSELV 1142
Db 1313 KTASPAIVSTATINGTTEYTTWCPIST--TESRQOQTLTVTVTSCSGVGCSETASPAI 1370
QY 1143 ATT-----QATNENGKSPSTDLTSSLTGCT-SASTSANSSELVTSQSV 1184
Db 1371 VSTATATVNDVVYVPTWPTQANEESVSSKMSNSATGETTTTLAAETTTNVAETITN 1430
QY 1185 TGA-----VASANDQSHSTSV-----TNSNSIVSNTPQTLSSQ 1220
Db 1431 TGAATKTVTSSLSRSNHAETQATSDVIGHSSSVSVSETGNTKSLTSSGLSTMSQ 1490
QY 1221 VTSSSP-----STNTFASTYDGSIIQHSHTWLYGLITLLSLFT 1260
Db 1491 PRSTPASSVMGVSTASLEISTYAGSANSLLAGSLGVSTFASULLAI 1536

RESULT 2

US-09-801-368-114
; Sequence 114, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Calli, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 114
; LENGTH: 1322
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-114

Query Match 14.4%; Score 932.5; DB 10; Length 1322;
Best Local Similarity 28.1%; Pred. No. 1.9e-37;
Matches 400; Conservative 215; Mismatches 468; Indels 339; Gaps 79;
QY 28 NSLTSNANAFKPGPPTWNAVIGWLDGTSANGDFTTLMPCVFKYTTSTQ-----S 83
Db 50 DSSYTSNAYMAY---GYASKT-----KLGSGVGGQTDISIDYNIPCV---SSSGTFFPCPQ 98

Db 1284 YET 1286

RESULT 8

US-09-815-242-12713

Sequence 12713, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 12713

LENGTH: 2344

TYPE: PRN

ORGANISM: Staphylococcus aureus

US-09-815-242-12713

Query Match 9.9%; Score 640.5; DB 10; Length 2344;

Best Local Similarity 22.2%; Pred. No. 4.2e-23;

Matches 323; Conservative 269; Mismatches 542; Indels 323; Gaps 46;

QY 17 AKTITGVDFSNLTWSNANAYAFKPGPYPTWNAVIGWSLDGTSANPGDFTFLNMPGVFK 76

Db 297 AITLGRIDSNKSFHFGKYNLNGYEGHNGGIGGFAP-----SPG----- 339

QY 77 YTTTSQTSVDLTADGVAYATCQFYSGBEFTFTSLTCTVNDALKSSIK-----AFGTV 128

Db 340 -VIGETGLNGAAVIGGLSNF--GPKLDYHN-TSTPNSSAKAKADPNNAVGGAGFAP. 395

QY 129 TLPFAFNVGGTSGSTOLE--SKCFAGTAGNTVTFNDGDKDISIDVEFEKSTVDPSPAYLYA 186

Db 396 VTTDSYGVASTYSSSTADNAALNVQPN--TFQD-----FDINYNGDT----- 439

QY 187 SRVMPSLNKVTLFLVAPQ-----CENGYSGTMTGTFSSNGSDVAIDCSNIHIGITK 236

Db 440 -----KVMYKYGAGQWTRISDWIAKSGTTFNFSLSMTASTGG-ATNLQVQVFG--- 487

QY 237 GLNDWNPVPSSESFSYTKTCTSNIGIOIKYONVAG-----YRPFIDAYISATDVNQY 288

Db 488 -----TFEYESAVT---QVRVYDVTGDKDIIPPKYSGNVQDVVTDNQQA 532

QY 289 TLAYNDYTCAGSRLOS-----KPFLLRWI-----GYKNSDAGS--- 322

Db 533 LTAKGNYTSSVDSYASTYNDNTNKTVMNAGOSVYYFTDVKAPTVTYVGNQIIEVGKIM 592

QY 323 NGVIVATFTVTDSTAVTTLFPNPSVDKTKTIEILOPI-----PT 364

Db 1284 YET 1286

593 NPIVLATTTONGTCTVTNTVTGLPSGLSYDSATNSIICTPTKIGQSTVTVVSTDOANKST 652

QY 365 TTITTSYGVGVTTSYLTAKTAPIGETATVI----- 392

Db 653 TTFITNVVDTAPTVT---PIGKSSEVFSPISPINIAODNSGNVNTVTVGLPSGLTF 709

QY 393 -----VDVPYHTTTTSEWTGTTTTT-----TRNPTDSIDTWWVQVPLPNP 436

Db 710 DSTNNTISGPTTNGITSTIIVSTDASGNKTTTTTFXYEVTNRMSDSVST--SGSTQSQ 767

QY 437 TVSTTEYWSQSFATT-----TTVAPPGGTDTVIIRE--PPNHTVTTTTEYWSQSFATT 488

Db 768 SVSTSKADQSASTSTSGSLMTSTASTSKSTSVLSDSVSASKLSSTESNSVSSSTST 827

QY 489 TTVTAPGGTDSVLIIREPPNPVTVTTEYWSQSFATTTVPAPPGGTDSVLIIREPPNPVT 548

Db 828 SLVNSQSVSS--MSGSVSKSTSLDFISNSSSTESKESVSTSTSDSLRSTSLSDSVSM 885

QY 549 TEYWSQSYATTTTVPAPPGGTDSVLIIREPPNPHTVTTEYWSQSYATTTTVP-- 601

Db 886 STSGSLKSKQSLSTSTSDASTSQSVSDSTNSISSTESLSESGSTSESISNSISNV 945

QY 602 -----PGTDTVIIREPPNPHTVTTEYWSQSFATTTVPAPPGGTDTVIIREPPNP 654

Db 946 SASTSKLESQSTISLSTSDSKMSTSESLDSTSTSDSVGSLVAGSQSVSTSTSDSM 1005

QY 655 TTTEYWSQSYATTTTVPAPPGGTDTVIIREPPNPHTVTTEYWSQSYATTT-----TTVT 707

Db 1006 STSEMIDSMSTSGSLAASDSKMSVSSSMSTSQSGSTSESLDSTSTSDSKSLST 1065

QY 708 APPGETDTVL-----IREPPNPHTVTTEYWSQSYATTTTVPAPPGGTDTVIIREPPNP 762

Db 1066 SQSGSTSTSTSTSSVMSSESQSTSGSMSTSQSDSTSTSTSTSTSTSTSTSTSTST 1119

QY 763 TTTEYWSQSFATTTVPAPPGGTDTVIIREPPNPHTVTTEYWSQSFATTSTSTSTSTST 822

Db 1120 ASSESISQSVST 1174

QY 823 TSDLSSTFESSMNTPTIS-----SDGMLLSSTTLVTE-----SETTELICSDGKE 869

Db 1175 TSD-SISEAISGSESTISLSESNSTSDSEKSAFLSESLSESTSESLSGSTSD 1233

QY 870 CSRLS-----SSSGIVTNPDNESSIVTSTVPTAST-----MSDSLSS 907

Db 1234 STSLSDSNSESGSTSTSLNSSTSGSASISTSTSGSASTSTVKSESVSTSTSTSTST 1293

QY 908 TDGISTATSDNVKS-----GVSVTTTSTVTTTQTPNPPLSSSVSTLTQLSSIPSESE 962

Db 1294 STSLSTSLDSTSGSKNSLSASMTSDSTSTTRKSESLASTSLSGSTSESGSTSSSE 1353

QY 963 SKVTFTS-----NGDNQSGTHD--SQSTSTEIIVTTSKVLPPVSSNTDITSE 1011

Db 1354 SKSDSTSMLSMSQSGSTSVSTSESLDSTSTSLASMAQSGVDNSASOSASTST 1413

QY 1012 PTNIRE--QPTTLTSTNSITDITTSQPT--GDNGDNTSTNPVPTVATSTLASAED 1067

Db 1414 STSTSEDSQSTSSYSTSQSTSQSESTSTSTSLSDSTSTSTSTSTSTSTSTSTSTST 1473

QY 1068 NKSGS-----HESASTSLKPSMG-ENSGLTSTTEIETATTSPTAPSPVSGCTDV 1117

Db 1474 SDSQSISTSTSEKSESTSTSLSDSTSTNSGSGASTSTLLNSASASESDSSSTSL-SDS 1532

QY 1118 TTEPTTRQPTTLTSTSTNSSELVAT-----TOATNENGKSPSDTLTSSLT 1165

Db 1533 TSASMOSSSESDSTSTSLNSQSTSTSTIRMTIASVESSESTSESGSTSESDSTST 1592

QY 1166 TGTSTASTSANSELVTSVGTGGAVASANDQSHSTSVTNSIVSNTPQTTLQQVTSST 1225

Db 1593 TSLSDSQST 1644

QY 1226 PSTNTFIATSYDGGSI 1242

Db 1645 TSTSTSVSDSTSLSDSV 1661

Db 598 -LSEASTTYSRSPRTTTLSPASMTSLGVE-ESTSRSPQSGTHSTVSPASTT---TP 652
QY 1001 VVS--SNTDLTSEPNTNR-----EQPTTL-----STSNISITEDITS--- 1036
Db 653 GLSEESTVYSSPGSTETVPRSTTTSVRGEETTFHSRPASTHTTLFTEDSTSGLT 712
QY 1037 -OPTGNGDNTSSTNPVPVATSTLASAEENK-SGSHESASTSLKPSMGNSGL----- 1090
Db 713 EESTAFPGSPASTOTGLP--ATLTTADLGEESTTFPSSSGSTGTLTSPARSTTSGLVGES 770
QY 1091 -----TTSTETETATTT-----SPTAPSPAVSSGTDVTEPTDTR 1125
Db 771 TPRSLSPSSTETTLPGSTPTSLSEKSTFTVTSRSPDATLSPATTTSGVSEESTSH 830
QY 1126 EQPTTLSTTS-----KTNSELVATTQATNENGCKSPSTDLTSSL----- 1164
Db 831 SQPGSTHTTAFDPSTTSLGSLQEPKTSHSSQSGTEAT-----LSPGSTTASLGGQSTTF 885
QY 1165 --TTGTSASTANSELVTSGSVTGGAVASANDQSHSTSVTNSNIVSNTPTTLTQQVT 1222
Db 886 HSPGDTETTLDPDDTITSLGVEASTPHTSSSTGSLHTTLTPASSTAGLQEEESTTFQSWP 945
QY 1223 SSSPST 1228
Db 946 SSSDPT 951

RESULT 12

US-10-142-515-11
; Sequence 11, Application US/10142515
; Publication No. US20030078399A1
; GENERAL INFORMATION:
; APPLICANT: SLOAN-KETERING INSTITUTE FOR CANCER RESEARCH
; APPLICANT: Lloyd, Kenneth O.
; APPLICANT: Yin, Beatrice W.T.
; TITLE OF INVENTION: Nucleic Acid Sequence Encoding Ovarian Antigen, Cal25, and Uses T
; FILE REFERENCE: 649-A-US
; CURRENT APPLICATION NUMBER: US/10/142,515
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/290,480
; PRIOR FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 5877
; TYPE: PRT
; ORGANISM: Human Being
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(5877)
; OTHER INFORMATION: Amino acid sequence of MUC16B
US-10-142-515-11

Query Match 8.0%; Score 522.5; DB 9; Length 5877;
Best Local Similarity 22.5%; Pred. No. 5.5e-17;
Matches 336; Conservative 211; Mismatches 568; Indels 379; Gaps 65;
QY 58 GTSNPGDFTLNMPCVFKYTTTSQTSVDLTADGVKATCQFYSGEEFTTSLTCVNDA 117
Db 1861 GITSPPESRFTMSV-----TESTHHLSTDLTP-----SAETIS-----TGTVMPS 1900
QY 118 LKSSIKAFGTVLPIAFNMGVGTGSDTDLSDSKCFAGTNTVTFNODGDKDISIDVEPEKST 177
Db 1901 LSEAMTSFATGVPRA--ISGSGSPFSRTESGPGDALTSLIA---ESLPSSTVPVPSST 1955
QY 178 VDPAYLYASRVNPSLUNKVTTTLFVAPQ-----CENGVTSG--TWGFSSSNGDVAIDCS 228
Db 1956 FTTT-----DSSITPALHEITSSSATPYRYDTSILGTESSTEGRLVMGTESST-----T 2004
QY 229 NIHIGITKGLNDWNPVSSSEFSYKTCSTNGIIQIKYQNPAGYR-PFIDAYISATD-VN 286
Db 2005 EGRLVNVTLDYSSQPGRTSSSPLDTRMTESVEL--GTVTAYQVPSLSTRLTRDGM 2062

QY 287 QYTLAYNDYTCAG-----SRLQSKPFTLRWTGYKNSDAGSNGIVIVATRTVTD 336
Db 2063 EHITKIPNEAAHRTGIRPVKGQPTSTSPASPKGLHTGGTKRMETTTTALKTTTALKTTTS 2122
QY 337 SITAVTTL-----PENPSVDKTKTIE-----ILQIPITTT--ITTSYGVGVTT 376
Db 2123 RATLTTSVYPTLGLTLPLNASOMASTIPTKEMMITPYVFPDVPETTSLATSLGAETS 2182
QY 377 SVLTKTAPICETATVVDVPHYTTTWTIS-----EWGTGTTTTRTNPTDSDIVVVQVP 432
Db 2183 TALPRTP-----SVFNRESEETASLVSRSGAERSPVQTLDVSSSEPDTTASWHPA 2236
QY 433 LPNPVTS-TTEYWSQSFAATTTTAPPGGTD-----TVIIREPPNHTV 474
Db 2237 ETIPTVSKTTPNFHSELDTVSSTATSHGADVSSAIPNTNISPSELDALTLPLVTSIGTDS 2296
QY 475 TTEYWSQS---PATTTTVPAPPGGTDVLIIRPPN-----PTVTTTEYWSQSFSAT 522
Db 2297 TTFPTLTKSPHETETRTTTLTHPAETSSTIPRTPNFHSHESDATPSIATSP-GAETSSA 2355
QY 523 TTTVTAPPGGTDVSI-----REPPNPTVTTEYWSQSAYATTTTVPAPPGTDSVII 574
Db 2356 IPIMTVPGAEDLVTQVTSSTGDRNMTPTLTLSPP--GEPKTIASLVTHHPEAQTSSAI- 2412
QY 575 REPPNHTV-----TTEYWSQSAYATTT---VTAPPGGTDTVI-----IREPPNHTV 618
Db 2413 ---PTSTISPAVSRSLVTSMTSLAAKTSITNRLTNSPGEPATVSLVTHSAQTSPTVPW 2469
QY 619 TTEYWSQSFAATTTVTGPPSGTDVLIIRPPNPTVTT-----TEYWSQSAYATTT--- 669
Db 2470 TTSIFFHKSDDTTPSMTTSHGAESSAV---PTPTVSTEVPGVVTPLVTSRRAVISTTIP 2526
QY 670 -ITAPPGETDVLIREPPNHTVTTEYWSQSAYATTTTVPAPPGTDTVLIREPPNHTVTT 728
Db 2527 ILTSLGPEPT-----PS---MATSHGEEASAIPTPTVSPGVGVTSVLTSSRAVTS 2578
QY 729 TEY-----WSQSAYATTTTVPAPPGTDT-----VIIREPPNPTVTTEYWSQSFAATTT--T 777
Db 2579 TTIPTLTFSLGPEPTTPSMATSHGTEAGSAVPTVLPVPCG-MVTSLVASSRAVTSSTLPT 2637
QY 778 VTAPPGGTDVLIIEKSSSSKISTSSNDITSIPTSPSRPHYVNSTSDLSSTFSSSSM--- 834
Db 2638 LTLSPGPEPT---PSMATSHGAEASSTVPTVSPV--PGVVTSLVTSSTSSGVNSTSIPTL 2692
QY 835 -----NPTTSISDGMLLSS-----TTLVT----- 854
Db 2693 ILSPGELETPSMATSHGAEASAVPTPTVSPGVGVVTPLVTSRAVTSSTTIPILTSS 2752
QY 855 -ESETTELICSDGKECSR--LSSS---SGIVTNPDNNESSIVTSTVPTASTMSDLSST 908
Db 2753 SEPETTPSMATSHGVEASSAVLVSPVPCGMVTFVLTSSRAVTSSTTIPLTISSDEPETT 2812
QY 909 DGLSATSSDNVSKS---GVSVTETSTVTIQTTPNPLSSSVTSLSLTQLSIPSV----- 958
Db 2813 TSLVTHSEAKMISAIPTLGVSPVQGLVTSVTSVSSGSETSAFNSLTAVASSOPEITDSWA 2872
QY 959 ---SESEKV-TFT-SNGDNQSG---THDSQSTST-----E 986
Db 2873 HPGTEASSVVTLVTSTGEPTNLSLVTHPAEASSLPRTTSRSHSELDTMPTSTVTSPE 2932
QY 987 IEIVTTSSTKVLPP-----VVSNDLTSEPTNTRQOFTTLSTTSNITEDITSQP 1038
Db 2933 AESSAISITISPGIVLTSVLTSSGRDISATFPTVPESHESEATASWVTHPAVISTT 2992
QY 1039 TGDGNGDNTSTNP--VPTVATSTLASA-----SEEDNKSGHESAS-T 1078
Db 2993 VPRPTPNYSHSEPTTPSIATSPGAETSDFPTITVSPDVPDMVTQVTSSTGDTSTIIP 3052
QY 1079 SLKPSMGE--NSGLTSTTEIATTTSPTEASP-----AVSSGTDVTT---EPTD 1123
Db 3053 TFLSSGEPEPTTSTFYTSTHTSSAIPTLPVSPADSKMLTSLVTSIGDSTTFFPTLTE 3112
QY 1124 TREQPTT---LSTTSKTNSELVATTQ-----ATNENGKSPS----- 1157

Db 3113 TPYEPETAIQLIHPAETNTMVRTPKFKSHSKSDTTLPAVITSPGPEASSAVSTTTISP 3172
Qy 1158 --TDLSSL--TTGTSASTS-----ANSELVTSGSVTGAVASANDQSHSTS 1201
Db 3173 DMSDLVTSLVPSFGTDTSTFFPLSETPYEPETATWLTHTPAETSTTVSGTIPNFHSHRS 3232
Qy 1202 VTNSNSIV-----SNTPOTLL-----SQQVTSSTSTNTFIATSYDGS 1240
Db 3233 DTAPSMVTSRGVDRSGVPTTTPPSIPGVVTSQVTSATDTSTAIPTLTPSPG 3286

RESULT 13

US-09-801-368-106
; Sequence 106, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 106
; LENGTH: 1169
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-106

Query Match 7.7%; Score 501; DB 10; Length 1169;
Best Local Similarity 23.1%; Pred. No. 1e-16;
Matches 313; Conservative 208; Mismatches 528; Indels 308; Gaps 53;
Qy 6 TLLFLVLSIASAKTIT-----GVFDSFNLSLWNSAANYAFKGPYPTWNAVILG 53
Db 11 TGLFL-LSVANVALGTEACLPAGEKNGMTINFYQYSLKDSSTYS--NPSYMAYGADA 67
Qy 54 WSLDGTSGANDPFTLNMPCVFKYTSQTSVDLTADGVKATQCFYSGE-----EFTT 106
Db 68 EKLGSVSGQTKLSDIYSPICNGASDTCACSD--DATEYSAQVVPVVRGVKLCSDNTT 124
Qy 107 FSTLT-----CTVNDALKSSIKAFKGTVTLPTAFNVGGTGSSTDLSDSKCFAGTNTVT 159
Db 125 LSSKTEKRENDCCQGAAYWSS--DLFGFYTPNTVMTG-----YFLPPKGTGT 175
Qy 160 FNDGDKDISIDVEFEKSTVDPSAYLYASRVMSLNKVTTLFVAPOCENGYSCTMGFSSS 219
Db 176 FG-----FATVDDSAI-----SV 199
Qy 220 NGDVAIDCSNIH---IGIT-----KGLNDWN--YPVSSSEPSYTKTCTNSGIIQYON-VP 269
Db 190 GGNVAFECCKQEQPPITSTDFTINGIKPNADAPTDLKSTYMYAGYYPYKIVYNAVS 249
Qy 270 AGYRPFIDAYISATDVNQ-----YTLATYND-----YTCAGSRLOSQPFLL 310
Db 250 WGTLPVSVVLPGDTEVNDDEGVVDFSDDNATQAHCSVPNPAEHARTCVSSATSS----- 304

RESULT 14

US-09-864-761-36047

; Sequence 36047, Application US/09864761

Qy 311 RTMGYKNSDAGSNGIVIVATRTVTDSTTAVTLPENPSVDKTKTIELQPIPTTTITS 370
Db 305 -WSSE-----VCTECTETESTSVTVTVYVSSSSWSSEVCTECTETESTSTP 352
Qy 371 YGVVTSYLTKT-APIGETATVIVDPYHHTTTVTSEWGTITTTTT---TRTPNTSIDT 426
Db 353 YVTSSSSSSEVCTECTETESTSVTVYVSSSSAAANYTSSFSSEVCTECTETEST 412
Qy 427 VVQVPLPNTVSTTEYWSQSFAATTVTAPPGGTDVVIIRPPNHVTVTEYWSQSFA 486
Db 413 STPIV-----TSSWSSEVCTECTE-----TESTSVTVYVSSSAAANYTSSFS 459
Qy 487 TTTVTAPPGGTDVVIIRPPNHVTTEYWSQSFAATTVTAPPGGTDVVIIRPPNPTV 546
Db 460 SSEVCTECTETES-----TSTPIYVTSSSSSSEVCTECTETESTSVTVYVSS 510
Qy 547 TTTEYWSQYATTTVTAPPGGTDVVIIRPPNHVTTEYWSQYATTTVTAPPGTD 606
Db 511 TAAANVTSSFSSEVCTECTETES-----TSTPIYVTSSSSSSEVCTECTETE 561
Qy 607 TVIIRPPNHVTTEYWSQSFAATTVTAPPGGTDVVIIRPPNHVTVTEYWSQYAT 666
Db 562 STSYTVYVSSSSTAAANYTSSFSSEVCTECTETES-----TSTPIYATSTGT 612
Qy 667 TTTITAPPGGTDVVIIRPPNHVTTEYWSQYATTTVTAPPGTDVVIIRPPNHVT 726
Db 613 ATSFATSTNTMTSLVQ-----TDTVSELSSTVSEHTNAP-----TSSVESNASTFI 661
Qy 727 TTTEYWSQYATTTVTAPPGGTDVVIIRPPNHVT-----TTEYWSQYATTTVTAP 781
Db 662 SSNGKGVSVYTSIHSITP-----MYPSNQVTSSSVSTPITSSSESSASVIL 713
Qy 782 PGGTDVVIYESSSSKISTSSNDITSIIPSF---SRPHYVNSTTSDLTSESSMNTPT 838
Db 714 P-STITSEFKPMTKTVKVSISSTNLITSIYDTSKDTSGVSTSVLSLISLPSY 772
Qy 839 SISSD---GMLLSSTTLVTESETTTELICDCKECSRLSSSGIVTNPDSNESSIVTST 894
Db 773 SASSEQIFHSSIVSSNGQALTSFSTKVSSESESHRTSPIT-----SSESGKSSG 825
Qy 895 VPTASTMSDLSSTDGISATSDNVSKSGSVVTTETSVTTIOTTPNPLSSVSTLTQLSS 954
Db 826 VEIESTSTSFSPHE--TSTASTSVQISSQFVTPSPISVA---PRSTGLNSQTE-- 876
Qy 955 IPSVSESEKVTFTSNGDNQSGTHDSQSTSEIETVTSKVLPPVWSSNTD----- 1007
Db 877 ----STNSKRETMSS-NSASVMPSSSATSPKTKVTSDET-----SSGFSRDRPTVY 924
Qy 1008 -LTSEPTNTREQPTTLSTT-----SNSITEDITTSQPTGNDGNTSSTNPVPTVATST 1059
Db 925 RMTSEPTNEQPTTLITVSSCESNSCNTVSSAVSTAATTINGITTEYTWCLPLSATE- 983
Qy 1060 LASASEDNKSGSHESASTSLKPSMGENGLTSTETIEATTSPTTAP---SPAVSSGTD 1116
Db 984 LTTVSKLESE-----EKTTLITVSCSGVCEATASPAIVSTATATVND 1027
Qy 1117 VTTEPTDTRQPTT-LSTTSKTNSEL-----VATTOATNENGKSPSTDLTSSIT-T 1166
Db 1028 VTVYVTSFQATNKLAIVSDIENSASKASFVSEAAETKISRNNFVPTSGTSTTETH 1087
Qy 1167 GTSASTSANSELVTSQSVTGAVASANDQSHSTSVTNSNSI--VSNTPQTTLQQVTS 1225
Db 1088 TTTNSASNSDNV-----SASEAVS-SKSVTNPVLISSVQOQPRCTPASSMIGS- 1134
Qy 1226 PSTNTFIATSYDGSGLIQTWLYGLITL--LSLFI 1260
Db 1135 -STASLEMSYLG---IANH-----LLTNSGISIFI 1161

; PRIOR APPLICATION NUMBER: US 60/125164
 ; PRIOR FILING DATE: 1999-03-19
 ; NUMBER OF SEQ ID NOS: 388
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 109
 ; LENGTH: 1236
 ; TYPE: PRT
 ; ORGANISM: Streptococcus pneumoniae
 ; US-09-769-787-109

Query Match 6.68; Score 428; DB 9; Length 1236;
 Best Local Similarity 23.3%; Pred. No. 3.7e-13;
 Matches 217; Conservative 156; Mismatches 433; Indels 126; Gaps 26;

Qy	327	IVATRTVDTSTRAVTTLPENPVDKTKTEILQPIPTT-TITTSVGVTTSTLYLTAPI	385
Db	22	IYAAGAVISCTVATQKVFNTESAVLEKTVKTDALATNDTVVLGTISTNSASTLSA	81
Qy	386	GETATVIVDPYHTTITVTSEWGTITTTTTRNPTDSIDTVVQVPLPNPTVTTEYWS	445
Db	82	SESA-----STSASEASTSASTSASEASTSA---STSISASTVWGS	126
Qy	446	QSFATTTVTAPGGTDTVIIRPP--NHTVTTTEWSQSFATTTVTAPPGTDSV--I	501
Db	127	QT-AAATEATAKKVEED-----RKKPASDYVASVTNVLQSYAKRR-----	176
Qy	502	IREPPNPTVTTEWSQSFATTTVTAPPGTDSVLIIRPPNPTVTTEWSQSYATTTT	561
Db	177	LASIKNAV-----FSGNTIVGAPAINASLNI-----AKSET	209
Qy	562	VTAPPGTDSVLIIRPPNPTVTTEWSQSYATTTVTAPPGTDTVIIRPPNHTVT	620
Db	210	KVVTGEGVDSV-YRVPYIKLVNDGSKLTFTYTVYVNPKNLGNISSMRPGYSIYN	268
Qy	621	TEWSQSFATTTVTGPPGTDVLIIRPPNPTVTTEWSQSYATTTITAPGETDTV	680
Db	269	SGTSTQMLTLGSLDKPKSPKKNYITDKNGRQVL-----SYNTSTMTTQSGYT---	317
Qy	681	LIREPPNHTVTTEWSQSYATTTVTAPPGTDTVLIIRPPNHTVTTEWSQSYATTT	740
Db	318	-----WNGAQMNGFFAKKGYGLTSSWTVPTGTDTSTFTTP-----YAART	359
Qy	741	T-----VTAPPGTDTVLIIRPPNPTVTTEWSQSFATTTVTAPPGTDTVIIYESMSS	797
Db	360	DRIGINFYNGGKGV-----ESSTTSQLSQS--KSLSVASQASASASA---STSASA	407
Qy	798	KISTSSNDITSIIIPSRPHVYVNTSDLSFTFESSMNTPTSISDGMLLSSTLTVESE	857
Db	408	SASTSASASTSASASA-----STSASVASTSASASTSASA-----SASTSASESA	457
Qy	858	TTTELICDCKGECRLSSSSGIVTPNPDNESSIVTSTVPTASTMSDLSLSTDCISATSSD	917
Db	458	STSASASTSASASTSASASTSASASTSASESTASASTSASESTASASTSASASTSA	517
Qy	918	NVSKS-----GVSVTTSTSTVTTQTPNPLSSSVTLTQLSSIPSVSESESKVTFTSNGD	972
Db	518	SASASTSASGASTSTASASTSASASTSASASTSASASISASESTASASTSASASTSA	577
Qy	973	NQSGTHDSQSTSEIEIVTSTTKVLPPVYSSNNTDLTSEPTNTREQPTTLSTNSITED	1032
Db	578	STAS-ESASTSASASTSASASTSASASTSASASTSASESTASASTSASESTASASTS	632
Qy	1033	ITTSQPTGNDGNTSTNPNPTVTATLASASEEDNKGSHESASTSLKPSMGENSLTT	1092
Db	633	TSASASTSASASTSASASTSASASTSASVASTSASASAS--TSASASTSASESA--ST	689
Qy	1093	STIEATTPTEAPSPVSSGDVTTPEPTDREQPTTLSTTKTNSSELVATTQATNENG	1152
Db	690	SASASTSASASTSASASTSASASTSASESTASASTSASESTASASTSASASTSASA	749
Qy	1153	GKSPSDDLTSSTLTGTASTNSANSELVTSQSVTGGAVASNDQSHSTSVTNSNIVSWT	1212
Db	750	SASTSASGASTSTASASTSASASTSASASTSASASTSASESTASASTSASESTASAST	809

Qy 1213 PQTTLQQVTSSTSPSTNTFTASTYDGSGLIIQ 1244
 Db 810 SASEASTSASASTSASASTSASASTSASARQ 841

Search completed: June 11, 2003, 17:22:23
 Job time : 71 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:11:32 ; Search time 18 seconds
(without alignments)
2059.605 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLOQFTLLFLYLSIASAKTI.....SIHQHTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	940	14.5	1537	1 US-08-325-267A-2	Sequence 2, Appli
2	651	10.0	2137	4 US-09-134-001C-4463	Sequence 4463, Ap
3	516	7.9	894	3 US-08-362-525-22	Sequence 22, Appl
4	516	7.9	894	3 US-08-971-692-15	Sequence 15, Appl
5	488	7.5	1721	3 US-08-928-361B-6	Sequence 6, Appli
6	487	7.5	1721	3 US-08-700-651-5	Sequence 5, Appli
7	487	7.5	1837	3 US-08-928-361B-5	Sequence 5, Appli
8	468.5	7.2	862	1 US-08-325-267A-4	Sequence 4, Appli
9	404.5	6.2	2035	1 US-08-046-585-5	Sequence 5, Appli
10	404.5	6.2	2035	1 US-08-393-703-5	Sequence 5, Appli
11	404.5	6.2	2035	5 PCT-US93-11721-5	Sequence 5, Appli
12	392	6.0	1638	4 US-09-071-035-258	Sequence 258, App
13	392	6.0	1638	4 US-09-071-035-262	Sequence 262, App
14	392	6.0	1638	4 US-09-071-035-266	Sequence 266, App
15	364.5	5.6	1022	4 US-07-757-022B-84	Sequence 84, Appl
16	364.5	5.6	1320	4 US-07-757-022B-46	Sequence 46, Appl
17	364.5	5.6	1320	4 US-07-757-022B-60	Sequence 60, Appl
18	364.5	5.6	1361	4 US-07-757-022B-40	Sequence 40, Appl
19	362	5.6	1049	4 US-07-757-022B-58	Sequence 58, Appl
20	362	5.6	1140	4 US-07-757-022B-104	Sequence 104, App
21	362	5.6	1313	4 US-07-757-022B-142	Sequence 142, App
22	362	5.6	1314	4 US-07-757-022B-50	Sequence 50, Appl
23	362	5.6	1354	4 US-07-757-022B-48	Sequence 48, Appl
24	362	5.6	1363	4 US-07-757-022B-52	Sequence 52, Appl
25	362	5.6	1404	4 US-07-757-022B-2	Sequence 2, Appli
26	362	5.6	1404	4 US-07-757-022B-62	Sequence 62, Appl
27	361.5	5.6	1038	4 US-07-757-022B-74	Sequence 74, Appl

28	361.5	5.6	1270	4 US-07-757-022B-44	Sequence 44, Appl
29	361.5	5.6	1311	4 US-07-757-022B-42	Sequence 42, Appl
30	359.5	5.5	941	4 US-07-757-022B-14	Sequence 14, Appl
31	350	5.4	650	3 US-08-362-525-2	Sequence 2, Appli
32	323	5.0	907	3 US-08-783-774-2	Sequence 2, Appli
33	323	5.0	907	4 US-09-328-599A-1	Sequence 1, Appli
34	323	5.0	907	5 PCT-US95-04611A-19	Sequence 19, Appl
35	322	5.0	1481	2 US-08-616-844-40	Sequence 40, Appl
36	322	5.0	1481	2 US-08-599-654-40	Sequence 40, Appl
37	322	5.0	1481	3 US-08-944-868A-40	Sequence 40, Appl
38	322	5.0	1481	3 US-08-944-423A-40	Sequence 40, Appl
39	322	5.0	1481	3 US-08-944-496-40	Sequence 40, Appl
40	310.5	4.8	249	3 US-08-700-651-15	Sequence 15, Appl
41	310.5	4.8	249	3 US-08-928-361B-20	Sequence 20, Appl
42	304.5	4.7	878	4 US-09-556-706B-2	Sequence 2, Appli
43	304.5	4.7	2409	6 5180808-2	Patent No. 5180808
44	304	4.7	1203	4 US-09-351-200-2	Sequence 2, Appli
45	303	4.7	216	3 US-08-928-361B-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-325-267A-2
; Sequence 2, Application US/08325267A
; Patent No. 5585271
; GENERAL INFORMATION:
; APPLICANT: WATARI, JUNJI
; APPLICANT: TAKATA, YOSHIHIRO
; APPLICANT: OGAWA, MASAHIRO
; APPLICANT: PENTTILA, MERJA
; APPLICANT: ONNELA, MAIJA-LEENA
; APPLICANT: KERANEN, SIRKKA
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
; NUMBER OF SEQUENCES: 7
; TITLE OF INVENTION: CONTAINING THEM
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,267A
; FILING DATE: 18-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP PCT/JEP94/00290
; FILING DATE: 24-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 38871/1993
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1537 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-325-267A-2

Query Match 14.5%; Score 940; DB 1; Length 1537;
 Best Local Similarity 26.0%; Pred. No. 5.9e-46;
 Matches 417; Conservative 244; Mismatches 523; Indels 418; Gaps 85;

QY 2 LOQFTLLYLSTASAKT-----ITGVDFSF-----NSLTWSNAANYAF-----40
 DB 10 LAVFTLLAL-TSVASGATEACLPAGQKSGMINFYQYSLKDSSTYSNAYMAYGYASKT 68
 QY 41 -----KGGGYPTW-----NAVLGWSLD--G 58
 DB 69 KLGSGGGQDISIDYNIPCVSSSGTFCQDESYGNMGCKGACNSQGIAYWSTDLFG 128
 QY 59 TSANPGDTTLLNM-----PCVFKVTTQ-----TSVD 85
 DB 129 FYTTPNV-TLEMTGYELPQGYTFKATVDDDSAILSVGGATAPNCCAQQPPITSTN 187
 QY 86 LTADGVK-----YATCOFYSGEFTTFTLCTVNDALKSSIKAFGT-----VTLP 131
 DB 188 FTIDGKIPWGGSLPPNIEGTVMYAGYYPM-----KVYNSNAVSMGTLPISVTLTP 238
 QY 132 IAFNVGTCSSDLEDSKCFAGTAVTFTNDGDKDISIDVEKSTV-DPSAYLYASRYM 190
 DB 239 -----DGTIVSDDFEG-----YVYSFDD-----DLSQSNCTVDPDSNYA-VSTTT 277
 QY 191 PSLNKVTTLFVAPQCENGYSCTGMGFSSNGDVAIDCSNIHIGITKGLDNVYVSSEF 250
 DB 278 TITEPWGTFSTSTEMTIVGTNGVPTDETIVIRTPASTIITTEPNWSTFTSTST 337
 QY 251 SVTKTCTNGIOIKYONVAGYRPFIDAYISATDVNOYTLAYNDYTCAGSLQSKPFTL 310
 DB 338 ELTTVTGTNGVRDETI-----VIRTPATTATITTEPNWSTFTSTSTEL-----384
 QY 311 RWTGYKNSDAGSNGI---VIVATRVTDSTAVITL-PFNSVDKTKT-----355
 DB 385 -----TIVTGNGLPDETIIIVIRPTATTATTQPNWDTFTSTEMTIVGTNGL 438
 QY 356 -----IEILOPPTTIIT-----SVGVTTISYLTAKTA-----PIGETAVIVDVPYH 398
 DB 439 PTDETIIVIRPTATTATTQPNWDTFTSTSTEMTIVGTNGLPDET-IIIVIRPTT 497
 QY 399 TTT-TVTSEWGTIT-----TTTTRN-PTDSIDTVVQVPLPNPT-VSTTEYWSOS 447
 DB 498 ATTAMTTQPNWDTFTSTSTEMTIVGTNGLPDE-TIIVIRPTATTATTQPNWDT 556
 QY 448 EATT-----TIVTAPPG-GTD--TVIIREPNHT--VTTEYWSQFATT-----TIVTAPP 494
 DB 557 FSTSTEMTIVGTNGLPDETIIIVIRPTATTATTTEPNWSTFTSTELTIVTGTN 616
 QY 495 G-GTDS--VIREPNHT--VTTEYWSQFATT-----TIVTAPPG-GTDS--VIREPP 542
 DB 617 GLPTDETIIVIRPTATTATTQPNWDTFTSTSTEMTIVGTNGLPDETIIIVIRPT 676
 QY 543 NPT--VTTEYWSQFATT-----TIVTAPPG-GTDS--VIREPNHT--VTTEYWSOS 591
 DB 677 TATTAMTTQPNWDTFTSTSTEMTIVGTNGLPDETIIIVIRPTATTATTQPNWDT 736
 QY 592 YATT-----TIVTAPPG-GTD--TVIIREPNHTV--TTTEYWSQFATT-----TIVT---635
 DB 737 FSTSTEMTIVGTNGVPDEIVIRPTSEGLISTTTEPWTGTFTSTSTEMTIVGTN 796
 QY 636 GPPSGGDTVIREPNHT--TVTTEYWSQFATT-----TTITAP---PGETDVLIREPP 686
 DB 797 GQPTDETVIRPTSEGLVITTEPWTGTFTSTSTEMTIVGTNGVPDETDEIVIRPT 856
 QY 687 NHTV--TTTEYWSQFATT-----TIVTAPPG-TD--TVIIREPNHTV--TTTEYWSOS 735
 DB 857 SEGLISTTTEPWTGTFTSTSTEMTIVGTNGQPTDETIVIRPTSEGLISTTTEPWTGT 916
 QY 736 YATT-----TIVTAPPG-GTD--TVIIREPNHTV--TTTEYWSQFATT-----TIVTAPPGT-785
 DB 917 FSTSTEMTIVGTNGVPDETDEIVIRPTSEGLISTTTEPWTGTFTSTSTEMTIVGTN 976

QY 786 -----DTVIYBSMSSSKI-----STSSNDITSIIPSPSRP-----816
 DB 977 GQPTDETVIRPTSEGLISTTTEPWTGTFTSTSTEMTIVGTCTNQPDDETIVIRPT 1036
 QY 817 --HYVNSTSD-LSTFESSMNTPTSISSDGMLLSSTLLVTESETT-----TELICSDGKE 869
 DB 1037 SEGLVTTTTTEPWTGTFTSTSTEMTIVGTNGLPDETIVIRPTTAAISSLSSSSSGQI 1096
 QY 870 CSRLLSSSSGIVNP--DSNESSIVTSTVPTASTMSLSLSDGISTSDNVDKSGSVST 927
 DB 1097 TSSITSSRIIT--PFYPSNGTSVISSSVISSTSLFTSSPVISSSVISSTTSTISF 1155
 QY 928 TETSVTTIQTNPPLSSVSLTQLSSIPSVSESKVFTFTNGDNGQSTHDSQSTSTEI 987
 DB 1156 SSSSKSV--IPTSSSTSGSSESETSSAGVSSS-----SEISSESKSPYSSSS--L 1205
 QY 988 EIVTTSST-----KVLPPVYSSNDLTSEPTNTREOPTTLSTTS-----NSITEDIT 1035
 DB 1206 PLVTSATTSEATASLPPATT-----TKTSEQTLIVTIVTSCESHVCTESISPAVS 1256
 QY 1036 SQPTGDNQNTSTN--PVPVATSTLASASBEDNKGSHESASTSLKPSMGENSEGLTWS 1093
 DB 1257 TATVTVSGVTTETWCPISSTTETTKQTKGTTEQTTETTKQTVVTVISSCESDVCSKTAS 1316
 QY 1094 TEIEATTTSPTEAPSPAVSGTDVTTPEPTDREOPTTLSTT-----SKTNSSELVATT-1145
 DB 1317 PAIVSTSTATINGVTEYTTWCPISIT--TESRQOTTLIVTIVTSCESGVCSETASPAIVSTA 1374
 QY 1146 -----QATNENGKSPSTDLSLSTGT-SASTSANSSELVTSVGTGA 1188
 DB 1375 TATVNDVTVYTPWRPQTANESVSKMNSATGETTTNTLAAETTTNTAAETITNTGAA 1434
 QY 1189 -----VASANDQSHSTSV-----TNSNSIVNTPQTTLSQQVTS 1224
 DB 1435 ETKTVVTSLSRSHHAETQTASATDVIGHSSSVSVSETGNTKSLTSLTSLGSLTMSQQPRST 1494
 QY 1225 SP-----STNTFIASYDGSISIOHSTWLYGLITLLSLFI 1260
 DB 1495 PASSVMGVSTASLEISTYAGSANSLLAGSLSVFIASILLAI 1536

RESULT 2

US-09-134-001C-4463
 ; Sequence 4463, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
 ; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 4463
 ; LENGTH: 2137
 ; TYPE: PRT
 ; ORGANISM: Staphylococcus epidermidis
 ; US-09-134-001C-4463

Query Match 10.0%; Score 651; DB 4; Length 2137;
 Best Local Similarity 24.2%; Pred. No. 3.7e-29;
 Matches 314; Conservative 252; Mismatches 541; Indels 192; Gaps 37;
 QY 19 TITGV-----FD-SFNSLTWSNAANYAFKGPVPTNAVLGWSLDCGTSANPGDTFTLNMP 72
 DB 745 TVTGLPQGLKFDASTNSIV-----GTPQTGINTTITESTDAGNK-----785
 QY 73 CVFKYTTSTQTSVDLTADGVKYATCOFYSGEEFTTFTSTLTCTVNDALKSSIKAFGTVTLPI 132


```

QY 478 -----EYWSQS-FATTTTVPAGGTDVSIIRPPNPTVTTEY-----WSQSPA 521
Db 113 CSNSOGIAYWSTDLFGYTT-----PTNVTLEMTGYFLPPQSGSYTFKFA 157
QY 522 TT-TTVPAGGTDVSIIRPPNPTVTTEYWSQSYATTTTVPAGGTDVSIIRPPNH 580
Db 158 TVDSSAILSVGGATAFNCACCAQQPPITSNF-----TIDGKPMWGS-----LPPNI 204
QY 581 TVTTEYWSQSY-----ATTTTVPAGGTDVSIIRPPNPTVTTEY-----DTVI 609
Db 205 EGTVMYAGYVPMKVYNAVSWGTLPLISVLPDGTTVSDDFEGVYVFFDDLSQSNCT 264
QY 610 IREPPNH-----TVTTEYWSQSFAT-----TWTGCP-----PSGTDVSIIRPPNPTV--TT 656
Db 265 VPDPSNAYAVSTTTTPEPTGTFTSTEMTGTGTGNGVPTDETVIVIRTPTESEGLIST 324
QY 657 TEYWSQSYATT-----TTITAPGE-TD--TVLIIRPPNPTV--TTTEYWSQSYATT----- 703
Db 325 TEPWTGTFTSTEVTTITGNGQPTDETVIVIRTPTESEGLISTTTTEPTGTFTSTEM 384
QY 704 TTVTAPPGTDTVLIIRPPNPTVTTEYWSQSYATTTTVPAGGTDVSIIRPPNH--T 761
Db 385 TIVTGNG-----OPTDEIV-----IVIRTPSEGLV 411
QY 762 VTTTEYWSQSFATTTTVPAGGTDVSIIRPPNPTVTTEY-----DTVIIE-----SMSSKISTSSNDITSIIP 811
Db 412 TTTTEPTGTFTSTEMTGTGTGNGVPTDETVIVIRTPTESEGLISTTTTEPTGTFTSTEM 471
QY 812 SFSR-----PHYVNSTSLSTFESSMNTPTSISSDGMLLSSTLTVESETTTTELICSDG 867
Db 472 S-SRPIITFPYPSNGTSVSSSVISSVSSLTSSPVI--SSSVISSSTTTSTISFSES 528
QY 868 KECSRLSSSGVITPNDSNESSIVTVPASTMSDLSSTGDISATSDNYSKGSVST 927
Db 529 SKSSVPTSSSTSGSESETSS-----AGSVSSSFISSESKSPYSSSLP 576
QY 928 TETSVTIOTNPLSSSVTLQSSIPSVSESEKVTFTNGDNQSGTHDSQSSTEI 987
Db 577 LVTSATTSOETASSLPATTKT-----SEQTLTVV-----TSCS 613
QY 988 EIVTTSKVLPPVSSNNTDLSEPTNTREQPTTLSTNSITEDITTQPTGDNQDNITS 1047
Db 614 HVCITES---ISPAIVSTATVYSGVTEYTWCPISSTE-----TTKQTKGT--EQIT 662
QY 1048 STNPVPTVATSLASEED--NKSGSHESASTSLKPSMGESGLT-----STEIEATT 1100
Db 663 ETTKQTKVVT---ISSCESDVCSKTASPAIVSTATVYVTPWTPOTANESVSSKMNSTATGET 719
QY 1101 TSPT-----EAPSPA-VSSGTD-----VTEPTDTREOPTTLSTSKTNSSELVAT 1144
Db 835 SSVSVSVERGNTKSLTSSGLTMSQQPRTPASSVMGYSTASLEISTAGSAT 887

```

RESULT 4

```

US-08-971-692-15
; Sequence 15, Application US/08971692
; Patent No. 6114147
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Immobilized proteins with specific binding
; NUMBER OF INVENTIONS: capacities and their use in processes and products.
; NUMBER OF SEQUENCES: 40
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,692
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 894 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-971-692-15

Query Match 7.9%; Score 516; DB 3; Length 894;
Best Local Similarity 26.3%; Pred. No. 6.3e-22;
Matches 251; Conservative 126; Mismatches 310; Indels 266; Gaps 43;

QY 437 TVTTEYWSQSFATTTTVPAGGTDVSIIRPPNPTVTTEY----- 477
Db 53 TYSNAAYMAYGASKTKLGSVGQTDISIDYNIPCVSSSGTTPCPOEDSYGNWGCKGMA 112
QY 478 -----EYWSQS-FATTTTVPAGGTDVSIIRPPNPTVTTEY-----WSQSPA 521
Db 113 CSNSOGIAYWSTDLFGYTT-----PTNVTLEMTGYFLPPQSGSYTFKFA 157
QY 522 TT-TTVPAGGTDVSIIRPPNPTVTTEYWSQSYATTTTVPAGGTDVSIIRPPNH 580
Db 158 TVDSSAILSVGGATAFNCACCAQQPPITSNF-----TIDGKPMWGS-----LPPNI 204
QY 581 TVTTEYWSQSY-----ATTTTVPAGGTDVSIIRPPNPTVTTEY-----DTVI 609
Db 205 EGTVMYAGYVPMKVYNAVSWGTLPLISVLPDGTTVSDDFEGVYVFFDDLSQSNCT 264
QY 610 IREPPNH-----TVTTEYWSQSFAT-----TWTGCP-----PSGTDVSIIRPPNPTV--TT 656
Db 265 VPDPSNAYAVSTTTTPEPTGTFTSTEMTGTGTGNGVPTDETVIVIRTPTESEGLIST 324
QY 657 TEYWSQSYATT-----TTITAPGE-TD--TVLIIRPPNPTV--TTTEYWSQSYATT----- 703
Db 325 TEPWTGTFTSTEVTTITGNGQPTDETVIVIRTPTESEGLISTTTTEPTGTFTSTEM 384
QY 704 TTVTAPPGTDTVLIIRPPNPTVTTEYWSQSYATTTTVPAGGTDVSIIRPPNH--T 761
Db 385 TIVTGNG-----OPTDEIV-----IVIRTPSEGLV 411
QY 762 VTTTEYWSQSFATTTTVPAGGTDVSIIRPPNPTVTTEY-----DTVIIE-----SMSSKISTSSNDITSIIP 811
Db 412 TTTTEPTGTFTSTEMTGTGTGNGVPTDETVIVIRTPTESEGLISTTTTEPTGTFTSTEM 471
QY 812 SFSR-----PHYVNSTSLSTFESSMNTPTSISSDGMLLSSTLTVESETTTTELICSDG 867
Db 472 S-SRPIITFPYPSNGTSVSSSVISSVSSLTSSPVI--SSSVISSSTTTSTISFSES 528
QY 868 KECSRLSSSGVITPNDSNESSIVTVPASTMSDLSSTGDISATSDNYSKGSVST 927
Db 529 SKSSVPTSSSTSGSESETSS-----AGSVSSSFISSESKSPYSSSLP 576
QY 928 TETSVTIOTNPLSSSVTLQSSIPSVSESEKVTFTNGDNQSGTHDSQSSTEI 987
Db 577 LVTSATTSOETASSLPATTKT-----SEQTLTVV-----TSCS 613
QY 988 EIVTTSKVLPPVSSNNTDLSEPTNTREQPTTLSTNSITEDITTQPTGDNQDNITS 1047
Db 614 HVCITES---ISPAIVSTATVYSGVTEYTWCPISSTE-----TTKQTKGT--EQIT 662
QY 1048 STNPVPTVATSLASEED--NKSGSHESASTSLKPSMGESGLT-----STEIEATT 1100
Db 663 ETTKQTKVVT---ISSCESDVCSKTASPAIVSTATVYVTPWTPOTANESVSSKMNSTATGET 719
QY 1101 TSPT-----EAPSPA-VSSGTD-----VTEPTDTREOPTTLSTSKTNSSELVAT 1144
Db 720 TLVTVTSCEGVCSETASPAIVSTATVYVTPWTPOTANESVSSKMNSTATGET 779

```


Db 1166 SNTG-----NIINPE---TGKVIPLGSLGSLNPSF-----NTPQOT--DEIT 1204
QY 1223 SSSPSTNTFIASDYDGS-GSIIQHSWTL 1249
Db 1205 GRPVDVTGL--PYDPSTGEIIDPATKL 1230

RESULT 6
US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4 (HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
US-08-700-651-5

Query Match 7.5%; Score 487; DB 3; Length 1721;
Best Local Similarity 22.3%; Pred No. 7, le-20;
Matches 301; Conservative 129; Mismatches 471; Indels 446; Gaps 57;

QY 214 MGFSSNGDVAIDCSNIHIGIKGLNDWNPVSSSEFSYKTC-----TSNGIQI- 263
Db 19 LNFTSTGTGTTDSNMNPVSIISG--ERRDPNKOATISGRSCGKQGVISDSSTGFRVD 76
QY 264 -----KYONVAGYRPFIDAVIS---ATDVOXYLAVTNDVTCAGSRLOSPFTLR 311
Db 77 SITGLPTDYSNCP--FNPVTGNLSRSTGKTIPNTYAGVRSNET-----KTTEPSANT 129
QY 312 WFG-YKNSD-----AGSNGIVI----- 327
Db 130 YAGVRSNETKTEPSANTNLLVDPKINAPCNSENSEQGOIFDMGSKVIYPTKCVV 189
QY 328 -----VATRVTDSTAVTLPNPNVDKTKTIEILOPIPTTITTSYGVGVTYS 378
Db 190 KHTTT 249
QY 379 LKTAPIGTAVIVDVPVHTTTTTVTSEWTGTTTTTTTTTNTDSDIDVWVQVPLNPTV 438
Db 250 TTTKKPT 309
QY 439 STTEYWSQSFAITTTVAPPDGTAVIIREPPNHTVTTTEYWSQSFAITTTVAPPDGT 498
Db 310 TTTT-TTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTT-----TTTTTTTTTTT- 359
QY 499 SVIIREPPNPTVTTTEYWSQSFAITTTVAPPDGTDSVIREPPNPTVTTTEYWSQSFA 558
Db 360 -----TTTKKPTTTT-----AT 402
QY 559 TTTVAPPDGTDSVIRE-----PPNHT 581
Db 403 TTTTISE---TESVTKPDWCWLEKNGEACEAGATYGVIGKDGRIENGMAFTMPNDOT 459
QY 582 VTTTEYWSQSFAITTTV-----TAPP----- 602
Db 460 HVRFRFKVDGNTISVRCRKAGKLEFPDRSLDFTIPPVAGHNCSIIIVGSGDKIHW 519
QY 603 ---GGTDVIREP--PNH-----TVTTEY-----WSQSFAIT-----TTVTG 636

Db 520 SPYSGKDVSLISAPIQPSSELFNEVYCDCTAKYGAHSGYQTSADFVITTTAKPTTTTG 579
QY 637 PPSGTDVTLIIREPPNPTVTTTEYWSQSFAITTTIAPPGETDVLIREPPNHTVTTTEY 696
Db 580 APGQPTTTTSGPSKPTTTT-----TKATTTTTLNP-----IITTT----- 617
QY 697 SOSYATTTTVPAPPGETDVLIREPPNHTVTTTEYWSQSFAITTTVAPPDGTDTVIIRE 756
Db 618 -TQKPTTTTTKVPG-----KPIATTTT---LXPIVTTTTTKATTTTITTV---- 661
QY 757 PNPPTVTTTEYWSQSFAITTTVAPPDGT-----TDTVIIYESMSSSKISTS 802
Db 662 ---PTTTTT--KRDEMTTTTPLDIDGIEITPIPIEKMLDKYRMIYDVYNSGLLDSN 716
QY 803 SN-----DITSIIIP-----SFSRPH-YVNSTTSDLT 828
Db 717 DEPIPGOAGQIADTSNLFPVQTHKSTGLPDMVGLPDPKSGNLVHPYTTQTMGLSV 776
QY 829 FESSMNTPTSISD-----GMLLSTTLVTESETTELICSDGKCSRLSSSS 877
Db 777 SYLAASN--LIVDTDETGLPIDTITGLPDLVPSLIPNPETGELFDPIDSDIMN-GTIA 833
QY 878 GIVTNPDSNES--SIVTSTVPTASTM-----SDLSSTDGI-----SATSSDNVS 920
Db 834 GIVSGISASESLLSQKSALIDPATNMVVEFGGLLPATGVMIPGLPGSEQTFSPETE 893
QY 921 KSGVSVTTTETSVTTTQTTNPPLSSSVTSLTQSSSPSVSES-----ESKVTFTSN--GDNQ 974
Db 894 DGGI-----IPPEVAAAANADKFKLISPPSPESIPERKDKIDSISELMYDIE 940
QY 975 SGTSDSQ-----STSTEIIVT-----TSSTKVLPPVPPVSSNNTDLTS 1010
Db 941 SGRLICQYKRPICSIACDLNPKMTPTQDSVTGKPIDPTTGLPFNPP-----TGHLI 995
QY 1011 EPTNTRQPTTUS-----TTSNSITEDITTSOPTGD--NGDNTSSNPVPTVAIS---TL 1060
Db 996 NPTNNTMDSBPAGAYKAVSNGIKTDNVYGLPVGEITGLPKDPGSDIPFNSTTGLVDP 1055
QY 1061 AASEEDKSGSHESASTSLKPSMGNSGL---TTSTETEATTTSPTEAPSPAVSSGTDV 1117
Db 1056 STGKPINNSTAGIVSGKPLPIEDENGLPDPNLPIDGNLQNLVNPETNTVSGSTG 1115
QY 1118 TTEP-----TDTREQPTTLSTTSKTNSELVATTQATNENGKSPSFDLTSS 1163
Db 1116 TTKPKPGIPVNGGVVPDEEAKDQ-----ADKKGDLIVPPTNSIN---KDPVINTQYS 1166
QY 1164 LTTGTSASTSANSSELVTSVGTGGAVASANDQSHSTSVTNSNSIVSNPQTTLTSSQVTS 1223
Db 1167 NTTG---NIINPE---TKVIPGSLGSLNPSF-----NTPQOT--DEITG 1205
QY 1224 SSPSTNTFIASDYDGS-GSIIQHSWTL 1249
Db 1206 KPVDVTGL--PYDPSTGEIIDPATKL 1230

RESULT 7
US-08-928-361B-5
; Sequence 5, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA

; ZIP: 94306-1840
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/928.361B
 ; FILING DATE: 12-SEP-1997
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/026.062
 ; FILING DATE: 13-SEP-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Veiriv, Hana
 ; REGISTRATION NUMBER: 30,518
 ; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 650-324-1677
 ; TELEFAX: 650-324-1678
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1837 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-928-361B-5

Query Match 7.5%; Score 487; DB 3; Length 1837;
 Best Local Similarity 22.8%; Pred. No. 7.8e-20;
 Matches 316; Conservative 146; Mismatches 487; Gaps 61;

QY 136 VGGTSSDLEDSKCFCTAGTATVTFN-DGDKDISIDVEFEKST---VDP-SAYLY--ASR 188
 DB 69 IDTGKFKSPYTGKHADASTSSAYSAPFELDVS-GVPIETNRMRWDPVSLMLFDNSTG 127
 QY 189 VM--PSLNKAVTFLFAPQCENGYTGCTMGFSSNGDVAIDCSNIHIGITKGLNDWNPVS 246
 DB 128 VMYDNTNLSLEAGTAGIRSESCIVSELNFSTTGTFTDTSNWPVSTSG----- 178
 QY 247 SESFSYTKTCSNGIQKQYVNPAGYRPFIDAYISATDVNQYTLAYNDYTCAGSRLOSK 306
 DB 179 -----ELKDPNKQA-----TISGSKS--- 194
 QY 307 PFTLWTKYKNSDAGSNGIVVATRTVTDSTAVTT-----LPNPSVDKTKTTEILQP 361
 DB 195 -----CGWK-----OGYSIDSSGFRVDSITGLPTDYPNCPFPN-----VTGNLVS 237
 QY 362 IPTTTITTSYGVVTSYLTAKA-PIGETATVVD-----TISGSKS--- 394
 DB 238 STGKTIPIWYAGYRSNETKTEPSANTNELLVDPKINAPCNSENSEFOVQIFDMGSKVY 297
 QY 395 VPY-----HTTTVTSEWGTITTTTTTRNPTDSIDTVVQVPLPNTVSTTEYWSQS 447
 DB 298 IPYTKCVGVKHTTTTTTTT-TT 356
 QY 448 FATTITVAPGGTDVLIIRPPNHTVITTEYWSQSFAITTTVAPPGGTDSVLIIRPPN 507
 DB 357 TTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 401
 QY 508 PVTITTEYWSQSFAITTTVAPPGGTDSVLIIRPPNHTVITTEYWSQSFAITTTVAPP 567
 DB 402 TTTTTTTTTTTTTTTTTTKKPTTTTTTTT-----TTTTTTTTTTTTTTTTTTTTTKKPT 457
 QY 568 GTDSVLIIRPPNHTVITTEYWSQSFAITTTVAPPGGTDSVLIIRPPNHTVITTEYWSQS 627
 DB 458 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-----TTTTTKKPTTTTTTTTTTKK 511
 QY 628 FATTITVAPGGTDSVLIIRP-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 649
 DB 512 PTTTTTATTTTTTSETESVIKPDWCWLEKNGCEAKGATVYGVICKDRIENGMAFTMI 571

QY 650 PN-PTVTTEYWSQSFAITTTT-----TAPP----- 674
 DB 572 PNDPTHVRFRFKVDGNTISVRCCKGAGKLEFPDRSLDFTIPPVAGHNSCIIIVGSGG 631
 QY 675 -----GDTDLVLIIRPPNHTVITTEYWSQSFAITTT-----TTVT 707
 DB 632 KKHVSPYCKSDVSLISAP-----IQPCELFNEVYCDTCTAKYCAIHSGYQTSADFVTTTT 687
 QY 708 APPGGTDVLIIRPPNHTVITTEYWSQSFAITTTVAPPGGTDSVLIIRPPNHTVITTEY 767
 DB 688 AKPTTTTGAQGP-----TTTTGSPSKPTTTTT--KATTTTIL--NPIITTT-- 734
 QY 768 WSQSFAITTTVTA---PFGGTDVLIIRPPNHTVITTEYWSQSFAITTTT-----VPTTTT 783
 DB 735 -TQKPTTTTTPKPKPPPIATTTTTLKPIVTTTTTKATTTT-----VPTTTT 783
 QY 825 DLSTFESSMNTP-----TSSSDGMLSSLTLTVESETELTICSD-----GKECS 871
 DB 784 TTKRDEMTTTTPLPDIGIEITPIPIEKMLDKYTRMYDNGSLGLLSDNDEPPIPGSQAG 843
 QY 872 RLS-----SSGIVTNP-----DSNESSIV-TSTVPTASTMSDLSSTDGISA 913
 DB 844 QIADTSNLFPVQTHKSTGLPDPMVGLPDPKSGNLVHPYTNQTMGSLSVSLAAKNTV 903
 QY 914 TSSDNVSKSGSVTTTETS-----VTTIQTP-----NPLSSSVTSLTQLSSIPSVSESE 962
 DB 904 DTDETY---GLPIDTLTGYPDPVSLIPNPETGELFDPISDEIMNGTIAGIVSGISASE 960
 QY 963 SKVTFTSN-----GD-----NOSGTHDSOSTSTEIEIIVTTSKTKVL 998
 DB 961 SLLSOKSAPIDPATNMVGEFGGLLNLPATGMIPGSLGPEQTPFSPEIE-----DGGII 1015
 QY 999 PB-VVSSNTD--LTSEPTNTRQPTTLSTTSNITE---DITTSQTPG-----NGDNT 1046
 DB 1016 PPEVAAAANKFKLSIPSPVSEIPEKQKIDISLMLAYDIESGRIGOVSKRPPIGSA 1075
 QY 1047 SNTNPVPTVATSLASASEDNKSGSHESASL--KPSMGNSGLTSTTEIEATTTSP 1104
 DB 1076 GDLEIMKTPTQT-----DSVTGKPIDPTTGLPFPNPTGH--LINTNNTMDSFA 1125
 QY 1105 EAPSAVSSG--TD-----VTTEPTDTRQPTTLSTTSNSELV--ATQATN----- 1149
 DB 1126 GAYKAVSNGIKTDNVYGLPVDITGLPKDPVSDIPFNSTTGELVDSPCKPINNYTAGI 1185
 QY 1150 -----ENGKK-SPSTD-----TSSLTTG-TSASTSANSELVTS 1183
 DB 1186 VSGKGLPIEDENGLNLPIDGNNQLVNPETNSVSGSTSKPKPKPIPVNG- 1244
 QY 1184 VTGGAVA-SASNDQSHSTS---VTNSNSIVSNTPOTTLISOQVTSSTSTNTFIASTYDG 1238
 DB 1245 --GGVVPDEAKDQADKGLIVPTNSI-----NKPDPVTNTQYSNT--- 1285
 QY 1239 SCSIIHST 1247
 DB 1286 TGNIIINPET 1294

RESULT 8
 US-08-325-267A-4
 ; Sequence 4, Application US/08325267A
 ; Patent No. 5585271
 ; GENERAL INFORMATION:
 ; APPLICANT: WATARI, JUNJI
 ; APPLICANT: TAKATA, YOSHIHIRO
 ; APPLICANT: OGAWA, MASAHIRO
 ; APPLICANT: PENTTILA, MERJA
 ; APPLICANT: ONNELA, MAIJA-LEENA
 ; APPLICANT: KERANEN, SIRKKA
 ; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT

STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0; Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325,267A
FILING DATE: 18-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP PCT/JP94/00290
FILING DATE: 24-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 38871/1993
FILING DATE: 26-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 862 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-325-267A-4

Query Match 7.2%; Score 468.5; DB 1; Length 862;
Best Local Similarity 25.1%; Pred. No. 3.2e-19;
Matches 241; Conservative 120; Mismatches 313; Indels 285; Gaps 39;

QY 437 TVSTTEWSQSFAITTVTPPGTDTVIIRPPNHTVTTT-----477
DB 53 TYSNAYMAYGASKTGLSGVGTDTISIDYNIPCVSSSGTFCPPQEDSYGNMGCKGMA 112
QY 478 -----EYWSQS-FATTTTVTPPGTDSVIIRPPNHTVTTT-----WSQSPA 521
DB 113 CSNSQGIAYWSTDLFGYIT-----PTNLTMTGTLFPQTGSYTFKEA 157
QY 522 TT-TTVTPPGTDSVIIRPPNHTVTTTTEYWSQSATTTTVTPPGTDSVIIRPPNH 580
DB 158 TVDSDAILSQVGAFAFNCQAQQPPITSTNF-----TIDGIKPWGS-----LPPNI 204
QY 581 TVITTEWSQSY-----ATTITVTPPGT-----DTVI 609
DB 205 EGVYMYAGYYPKVVYSNVSWGLPISVTLPGDTVSDDFEGYVYFDDDLDSQSNCT 264
QY 610 IREPPNH-----TVITTEYWSQFATT-----TTVTGP-----PSGTDTVIIRPPNPT--VIT 656
DB 265 VPDPSNVAVSTTTTTEPWTGTSTSTMTTGTGTGTGTGTGTGTGTGTGTGTGTGT 324
QY 657 TEYWSQSATTTTITAPPGETDVLIREPPNHTVTTTTEYWSQSATTTTITAPPGETDV 716
DB 325 TEPWTGTGTSTST-----EMTTVTGTNG-----347
QY 717 LIREPPNHTTTTTEYWSQSATTTTITAPPGETDVLIREPPNPT--TVITTEWSQSFAT 774
DB 348 ---OPTDET-----IVIRPTSEGLVTTTTEPWTGTFTS 379
QY 775 TTVTAPPGET-----DTVIIE-----SMSSKISTSNNDITSIIIPFSR-----PHYVN 820
DB 380 TSTMTSVTGTNGLPDTDETIVVKTPTTAISSLSLSSSQITSSITS-SRPIITFPYPS 438
QY 821 STTSDLSSTFSSSMNTPTSISSDGMLLSTTLVTESETTELICSDGKCSRLLSSSGIV 880

DB 439 NGTSVISSSVISSVTSSTLFTSPVI--SSSVISSSTTTTSTSFSSSKSVIPTSSSTS 496
QY 881 TNPDSNESSIVTSTVPTASTMSDLSLSTDGISATSDSNVSKSGSVTTTETSVTTIQTTPN 940
DB 497 GSSESETSS-----AGSVSSSSSFISSESKSPYSSSLPLVTSATTSQFTAS 544
QY 941 PLSSSVTSLTQLSSIPSVSESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTTSTKVLPP 1000
DB 545 SLPATTTKT-----SEQTLTVV-----TSCSHVCTES---ISPA 578
QY 1001 VVSNTDLTSEPTNTREQPTTLSTTNSITEDITTSQPTGDNDNTSSNPVPTVATSTL 1060
DB 579 IVSTATVTVSGVTEVTWPCISTE-----TTQTKGTT-EQITETTKQTIVVT--- 627
QY 1061 ASASEED--NKSGSHESASTSLKPSMGENSEGLT-----STEATNTTSTP----- 1104
DB 628 ISSCEDVCSTASPAIVSTSTATINGVTTEYTWCPISTTESRQQTTLVTVTSCSGVC 687
QY 1105 -EAPSPA-VSSGTD-----VTTEPTDTREQPTTLSTTSTNSSELVATQATNENGKSPS 1157
DB 688 SETASPAIVSTATATVNDVVTVTPWRPQTANEESVSSKMSATGETT--TNTLAAETTT 745
QY 1158 TDLTSLTGTSTASTSANSSELVTSGVTGG-----AVASASNDQSHSTSV-----TNSNS 1207
DB 746 NTVAETIITNGA---AETKTVVTSLSRSNHAETOTASATDVIGHSSSVSVSETGNTK 802
QY 1208 IVSNTQTTLTSLQVTSSTSP-----STNTFIATSYDGSGLIOHSTWGLIITLLSLFI 1260
DB 803 SLTSSGLTMSQQPRSTPASSVMGYSTASLEISYAGSANSLLAGSLSVFIASLLAI 861

RESULT 9

US-08-046-585-5
; Sequence 5, Application US/08046585
; Patent No. 5453362
; GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Heir, Winship
; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
; TITLE OF INVENTION: HOST CELL FACTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBRACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/046,585
; FILING DATE: 12-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57503-1/RAO
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

Db 1324 T H E T G T T H A T T A T S N G G T G Q P E G G Q P P A G R P C E T H Q T T S T G T T M S V S V G A L L P D A T S S 1383

843 QY ---GMLSSLLVIVSEIIL : :::: || | | | | | | | | | | | | | | | | | | | | | |
1384 Db HRTVESGLVEAAAPSVTPQAGTALLAPPFTQRVCs-----NPKSCVTETTSVTTCITOTNPISL 943
QY 889 SVITSVPASTMSDLSIGSATSSD-----NVKSGVSVTETSTVTTICOTTNPISL 1426
Db 1427 ---TGTHATTATTVSNMSSNQDPPPAASDOGEVESTQGDSVNITSSSAITT--TVSSTLT 1481
QY 944 SSVTSUTQSIPSPSESKEKVFTFNGONQS---GTHDSGSTSTEI-----EIVTSTSSTK 996
Db 1482 RAVTTVVQTSPVPSPPEELQVSPGRQQLPPLRQLLQASATALMGESAIEVLASQTP 1541
QY 997 VLPPPVSNTDLTSETNTREOQPTTLTSTNSITEIDITTSOPGDNDGTSSNNPVPVTA 1056
Db 1542 EUPAAVDLSS--TGPSSGOE-----SAGSAVAATVVVOPI-----PPTOS 1580
QY .1057 TSTLASASEB---DNKGGSHESASTSLKPMSGENSGILTTSETEATTTSPTEAPSAPVSS 1113
Db 1581 EVDDLSLPQELMAEAQAAGITLMVTGLTP---BELAVTAAAEEAAAAATEEAQALAIQA 1637
QY 1114 GTDV-----TTEPTDTTRQPPTTLSTTKTNSELVATTOATHENGKSPTSOLDTSSLTT 1166
Db 1638 VLAQAQAQVNGTCEPMDEATAATV-----TQAELE----GHLSAEGEQGAATTPIVLTQ 1688
QY 1167 GTSASTSANSELVTGS-----VTGGAVASANDOSHSTSYTHSNSIVSNTPQTT-- 1216
Db 1689 QELAAILVQOOLOQEAOAQOHHHLPTALPADSLNDPAIESNCL--NELAGTVPSTVAL 1746
QY 1217 LSQOOTSSPSPTNTFFIA 1233
Db 1747 LPSTATESLAPNSTFFVA 1763

RESULT 10
US-08-393-703-5
; Sequence 5, Application US/08393703
; Patent No. 5585239
GENERAL INFORMATION:
; APPLICANT: Lamarco, Kelly
; APPLICANT: Wilson, Angus
; APPLICANT: Herr, Winship
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
TITLE OF INVENTION: HOST CELL FACTOR
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,703
FILING DATE: 24-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A.
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-57503-2/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2035 amino acids

; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 US-08-393-703-5

Query Match 6.28; Score 404.5; DB 1; Length 2035;
 Best Local Similarity 21.9%; Pred. No. 4.9e-15;
 Matches 319; Conservative 165; Mismatches 506; Indels 467; Gaps 69;

QY 14 IASAKITGVDFDSFNSLTHSNAANYAFKPGYPTWNAV-LGWSLDGTSN-NGCDTTLNM 71
 Db 537 IGSSPQSGM-----AALAAAAATKIPSPASPTVLSVAGTIVKMTAVTPG---TTTL 589
 QY 72 PCVEKVTTSQTSVDLTADGVKATCFYSGEETFTSTLCTVNDALKSKIKAFGVTLTP 131
 Db 590 PATVKVASSPMV-----SNPATRMLKTRAAQVCTSVSS---ATNTSTRP 631
 QY 132 IAFNVGSGTSDLEDSKCTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMP 191
 Db 632 I-ITVHKSGVTVAQAQVVTTVVGGVT-----KTITL-----VKSPISVPGGSALIS 678
 QY 192 SLANKVTTLFVAPQCENGYSGTMGFSSNGDVAIDCSNTHIGITKGLNDWNPVSSSESFS 251
 Db 679 NLGKMSW---QTKPVQTSAVTG-QASTGPVT-----QIIQTKG----- 714
 QY 252 YTKTCTSNIGIKYQNVNPGYRPFIDAYISATDVNOYTLAYTNDYTCAGSRLOSKPFTLR 311
 Db 715 -----PLPAG---TILKLVTSADGKPTTIITTTQASGAG---TKTIL- 751
 QY 312 WTGYKNSDAGSNIGIVATRTVDTSTTAVTTLFPNPSVDK-----TKTIELQIP-- 363
 Db 752 -----GISSVSPSTKPGTGTIIKTPMSAIIITQAGATGVTSSPGIKSPITII 799
 QY 364 TTTITTSYGVTTSYLT-----KTAPIGETATVIVDVPVHTTTVT 404
 Db 800 TTKVMTSGTAPAKIITVPKIKATGHGQOQGVTOVLKGP-GOPGTILRTVPMGVRLVT 858
 QY 405 SEWGTGTTTTTRNTDSDITVVQVPLNPTVSTTEYWSQSFATTTVTAPPGGDTV 464
 Db 859 -----PVTSAVKP-AVTLVKGTTGVTGLTGT-----GIVSTSLAGAGHST- 902
 QY 465 IIREPNHVTTEYWSQSFATTTVTAPPGGDTSDVIIREPNPTVTTEYWSQSFATTT 524
 Db 903 -----SASLATPITL---GTATLSSQVINPTALTWS-----AAQT 936
 QY 525 TVTAPPGTSDVIIREP-PNPTVTTEYWSQSFATTTVTAPPGGDTSDVIIREPNHTV- 582
 Db 937 TLTAAGLTPTTITMPVSOPT-----QVTLITAPSG-----VEAQPVHDL 978
 QY 583 -----TTTEYWSQSFATTTVTAPPG---GTDTVIIREPP---NHTVTTTEYWSQSFAT 630
 Db 979 VSILASPTTE--QPTATVTIADSGQDVQPGIVTLVCSNPPCETHETGTT-----NTAT 1030
 QY 631 TTTVT---GPPSG-----TDTV-----IIREPNPTVTTEYWSQSFAT 664
 Db 1031 TTVVANLGHGPQTQVQFVCDRQEAASLVTSTVGQNGSVVRVCSNPPCETHETGTTNT 1090
 QY 665 ATTTT-----ITAPPGET-----DVLIREPNH-----T 689
 Db 1091 ATTATSNMAGHGCNPPCETHETGTNTATTAMSVGANHQRACAAAGTPAVIRIS 1150
 QY 690 VTTTEYWS-----QSVATTTVT-----AP-----PGTDTVLIR 719
 Db 1151 VATGALEAAGSKSQCTROTTSATSTVMTATGAPCSAGPILGPNMAREPGRSFAFVQ 1210
 QY 720 EPP-----NHTVTTTEYWSQSFATTTVTAPPGGDTSDVIIREPNHTV 740
 Db 1211 LAPLSKVRLSPIKDLPAGRSHAVSTAAMTRSSVSGAGEPRMAPVCSLOGGSPSTTV 1270
 QY 741 TVTA-----PPGCTDVIIREPNPTVTTEYWSQSFATTTVTAPPGGDTSDVIIREPNHTV 795
 Db 1271 TVTALEALLCPSATVTOVCSNPPCET-----HETGTTNTATTSNAGSAQRVCSNPPCE 1323

QY 796 SSKITSSNDITSII-----PSFSRPHYVNVSTTSDLTSE-SSSMNTPTSISS 842
 Db 1324 THETGTHHTATTATSNCTGQPEGGQPPAGRCPTHQTTSTGTTMSVSGALLPATSS 1383
 QY 843 -----DCMLLSSTTLVTESETT-----TELICDGGKECRSLSSSSGIVTNP--DSNES 888
 Db 1384 HRTVESGLEVAAPSVTPQAGTALLAPFPTQVCS-----NPPCETHE- 1426
 QY 889 SIVTSIVTPASTMSDLSLSDTGISATSSD-----NVSKSGVSVTTTSTVTTIOTTNPLS 943
 Db 1427 ---TGTHHTATTATSNMSSNQDPPPAASDQGEVESTQGSVNITSSAITT---TVSSTLT 1481
 QY 944 SSVTLTSLSPISVSESESKVTFTTNGDNQS--GTHDSQSYSTEI-----EIVTTSSTK 996
 Db 1482 RAVTTVTQTPVPGSPVPPPEELQVSPGQQLPPRQLQASASTALMGESAELVSASQTP 1541
 QY 997 VLPVVSNTDLTSEPTNTREQPTTLTSTNSITDITTSQPTGDNQNTSSNPVPTVA 1056
 Db 1542 ELPAADVLS--TGEPSSGQE-----SAGSAVATVVVQP-----PPPTQS 1580
 QY 1057 TSTLASASEE---DNKSGSHESASTSLKPSMGENSEGLTTSLEATEATTSPTTEAPSAVSS 1113
 Db 1581 EVDQLSLPQELMAEQAAGTTLMTGLTP---EELAVTAAEAQAAQAAATEEAQALAO 1637
 QY 1114 GTDV-----TTEPTDTRREQPTTLTSTSKTNSSELVATTQATNENGGKSPSTDLTSSLT 1166
 Db 1638 VLQAAQAAQVMTGCEPMDTSEAAATV-----TQAEI---CHLSAEGQEGQATTIPVLQ 1688
 QY 1167 GTSASTSANSSELVTSQS-----VTGGAVASASNDQSHSTSVTNSIVSNTPOTT-- 1216
 Db 1689 QELAAALVQOQLOEQAQAOQHHLPTALAPADSLNDPAIESNCL--NELAGTVPSTVAL 1746
 QY 1217 LSQOVTSSSPSTNTETA 1233
 Db 1747 LPSTATESLAPSTNTFA 1763

RESULT 11

PCT-US93-11721-5
 ; Sequence 5, Application PC/TUS9311721
 ; GENERAL INFORMATION:
 ; APPLICANT: Lamarco, Kelly
 ; APPLICANT: Wilson, Angus
 ; APPLICANT: Heir, Winship
 ; TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
 ; TITLE OF INVENTION: HOST CELL FACTOR
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
 ; STREET: 4 Embarcadero Center, Suite 3400
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94111-4187
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/11721
 ; FILING DATE: 03-DEC-1993
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Osman, Richard A
 ; REGISTRATION NUMBER: 36,627
 ; REFERENCE/DOCKET NUMBER: FP-57503-1/RAO
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 781-1989
 ; TELEFAX: (415) 398-3249
 ; TELEX: 910 277299
 ; INFORMATION FOR SEQ ID NO: 5:

Query Match	6.2%	Score 404.5	DB 5	Length 2035
Best Local Similarity	21.9%	Pred. No. 4.9e-15		
Matches	319	Conservative 165	Mismatches 506	Indels 467
Gaps	69			
QY	14	IASAKITIGVDFDSFNSILTWSSNAANYAFKPGYPTWNAV-LGWSLDGTS-A-NPGDFTFLNM	71	
DB	537	IGSPQMSGM---AALAAAAATQKIPPSAAPTJVSVPAGTTIVTKMAVTPG---TTTL	589	
QY	72	PCVFYKTTTSQTSYDLTADGVKYATCOPYSGEEETFTSTLCTCTVNDALKSSIKAFGTVTL	131	
DB	590	PATVKVASSPMV-----SNPATRMLKTAQAQVGTSSV---ATNTSTRP	631	
QY	132	IARNVGGTSGSTDLDSKCFAGTNVTFNDGDKDISIDVEFEKSTVDPDSAYLYASRVMP	191	
DB	632	I-ITVHSGTVTAQAQVTVTVGGVT-----KTITL-----VKPSIVPGGSAIS	678	
QY	192	SLANKVTTFLVAPOCENCYTSGTGFSSSGNDVAIDCSNIHIGITKGLDNPNYPVSSSEFS	251	
DB	679	NLGRVMSV---QTKPVQTSVATG-QASTGPVT-----QIIQTKG-----	714	
QY	252	YTKCTSNGIQIKYQNPYAGRYPIDAYISATDVNQVTLAYTNDYTCAGSLQASKPFTLR	311	
DB	715	-----PLPAG---TILKLVTSADGKPTTIITTTQASGAG---TKPTIL-	751	
QY	312	WTGKNSDAGSGIVIVATRTVTDSTTAVTTLFPNPSVDK-----TKTIEILOQTP--	363	
DB	752	-----GISSVSPSTRKPGTTTIITKTPMSAITTQAGATGVTSSPGKSPITII	799	
QY	364	TTTITTSYGVGTTSYLT-----KTAPIGATATVIVDPVHTTTT	404	
DB	800	TTKVMFSGTCAPAKIITAVPKIATGHQCGQVTVQLKAP-GQPGTILRVPVPGVGRVLT	858	
QY	405	SEWGTITTTTTRTNPDSDITVVQVPLPNPTVSTTEYWSQSFAATTTVTAPPGGPTDV	464	
DB	859	-----PVTSAVKP-AVTLVVKCTGTGTTLGTVT-----GTVSTSLAGAGHST-	902	
QY	465	IIREPPNHTVTTEYWSQSFAATTTVTAPPGGTDSVIIREPPNPTVTTEYWSQSFAATTT	524	
DB	903	-----SASLATPITL-----GFIATLSSQVINPITATVS-----AAQT	936	
QY	525	TVTAPPGTDSVIREP-PNPTVTTEYWSQSATTTTVPAGGTDVSVIIREPPNHTV-	582	
DB	937	TLTAAGGLTPTITMQPVSQPT-----QVTLITAPSG-----VEAQPVHDL	978	
QY	583	-----TTTEYWSQSATTTTVTAPPG-----GTDVIREPP-NHPTVTTEYWSQSFA	630	
DB	979	VSILASPTTE--QPTATVTIADSGQDVGQGTVLCSNPPCETHETGTT-----NTAT	1030	
QY	631	TTTIVT---GPPSG-----TDTV-----IIREPPNPTVTTEYWSQS	664	
DB	1031	TTVVANLGGHPQTPQVQVCDROEAAASLVTSTVGQNGSVRVCVSNPPCETHETGTTNT	1090	
QY	665	ATTTT-----ITAPPGET-----DTVLIREPPNH-----T	689	
DB	1091	ATTATSNMAGCHCSNPPCETHETGTTNTATTAMSSVGNHQRDARACAGTPAVIRIS	1150	
QY	690	VTTEYWS-----OSYATTTT-----AP-----PGETDVLIR	719	
DB	1151	VATGALEAAQGSQCOTRQTSATSTMTVMATGAPCSAGPALLGPSMAREPGGRSPAFVQ	1210	
QY	720	EPP-----NHTVTTEYWSQS-----ATTT	740	
DB	1211	LAPLSSKVRUSSPSIKDLPAGRHSIAVSTAAMTRSSVAGCEPRMAPVCESLQOGGSPETT	1270	
QY	741	TVTA-----PPGGTDVIREPPNPTVTTEYWSQSFAATTTVTAPPGGTDVILYSSMS	795	

TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 258:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1638 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-071-035-258

Query Match 6.0%; Score 392; DB 4; Length 1638;
 Best Local Similarity 22.4%; Pred. No. 1.9e-14;
 Matches 311; Conservative 149; Mismatches 603; Indels 326; Gaps 60;

QY 13 SIASAKTIGVDFSNLSWNAANYAFKGPYTNNAVLT-----GWSLDGTSANPG-DT 66
 Db 351 SEVTATITGV-----TGNSTAGYEVKGT--DANATVEIRNAGCTVIGTGTADGCA 401
 QY 67 FTLNMPG-VFKYTTTSQTSVDLADGVKY--ATCQYSGEFTTFTLT-----CTVNDAL 118
 Db 402 FTVTVPAGEAGANETLTA VAKNASGTEXTPTTFQPADEATVTA PTTTGTGNSGAGYEV 461
 QY 119 KSSIAFGVTVPIAFNVTGSGSDLEDSKCFAGTNTVTFNDGDKDISIDVEREKSV 178
 Db 462 KGTADANATVEIR---NAGGTVIGGTADG---TGAFTVTVPAGEAGAN-----ETLTA 509
 QY 179 DPSAVLYASRVMPSLNKVTLFVAPQCENGYTSGTMGFSSNGDVAIDCSNIHIGITKGL 238
 Db 510 -----VAKNASGTESTPTTQTPADEATVTA PTTTGTGNSGAGYEV-----VKGT 554
 QY 239 NDWNPVS---SESFSYTKCTGNGIQKQYQNPAGYRPPIDAYISATDVNQVTLAYTNDY 296
 Db 555 ADANATVEIRNAGGAVIGTGTADGTAFTVTPAG-----EAGANETLTA VAKN- 603
 QY 297 TCAGSRLOSKPRLRWTVGKNDSAGSNGIVIVATFTVYDSTTAVTTLFPNFSVDTKTI 356
 Db 604 ---ASGTSTPTTFQ-----TPADPNTVATPIVETVGTSTKGYEVKGTAEVG--TTI 652
 QY 357 EILQIPPTTTITTSYGVVTSVLTKTAPIGETATVLDVVPYHTTTTTSVETGTTTITTT 416
 Db 653 EV-ROAGTVLTATGTDGKY-TVTLDSG-TATA-----NQLSVVAKNASGTSQPAT 704
 QY 417 RNPTD-----SIDTV-----VVQVPLNPVTSTTEYWSQSFAITTT-----TVTA 456
 Db 705 ATTPADVATPTVDNITGNSGSGYEITGTADPNTTIEVRDPGAVIGTGTSDANGDFTVL 764
 QY 457 P-----PGCTDVIILREPNNHVTTEYWSQSFAITTT-----IVTAPP 494
 Db 765 PTGTTNPGDTLIVIGKDNAGNESQPTFVLVPADATVTA PTVGTGNSVAGYQVGTGADP 824
 QY 495 GGTDSVILREPP-NPTVTTTEYWSQSFAITTTV-TAPPGGTSVILREPNNPTVTTTEY 552
 Db 825 NAT--IEIRDADGNVIATGADTGDGTSFAVNLPAAGTANANETLTA LAKDPAGNTSPTTFQ 882
 QY 553 SOSYATTTVAPPGTSVILREPNNHVTTEYWSQSFAITTTVAPPG---GTDVVI 609
 Db 883 TPA---DEWAPP-SVDKVTGNTTQGYQVGT---AELGTTIEVRADGTGVLGTAT--- 931
 QY 610 IREPNNHVTTEYWSQSFAITTTVTPGSPGSDTVII-----REPNNPTVTTTEYWS 661
 Db 932 -----TGCTGQY-----TVTLASGKATQTVNVVAKNDTGLESQTTTAMTAD--- 975
 QY 662 QSYATTTTI-----TAPPGTDTVILREPNNHVTTEYWSQSFAITTT-T 704
 Db 976 ---VTTPTIGDITGDSITGCTGADP--NTTIEVRNPDGTLIGTTTTDDOQNTVDLP 1030
 QY 705 TVTAPPGETDVIILREPNNHVTTEYWSQSFAITTT--TVTAPPGETDT---VIIREPN 759
 Db 1031 AGAANPGDTLVVVGKDGNGSQPTFVTPEDATVAAPTPTVTVGTGTTATGYQVGTGAEPN 1090
 QY 760 PTVTTTEYWSQSFAITTT-----TVTAPPGETDVIILREPNNHVTTEYWSQS 804
 Db 1091 VTIEHNEAGLVIATGTTDCAFTAFTLTLTGATANEALTA LAKDAAGKESNPTAFKTPA 1150

RESULT 13

US-09-071-035-262
 ; Sequence 262, Application US/09071035
 ; Patent No. 6448043
 ; GENERAL INFORMATION:
 ; APPLICANT: Gil H. Choi
 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 ; NUMBER OF SEQUENCES: 496
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/071,035
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: A. Anders Brookes
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB369P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 262:

QY 805 --DITSIIIPSFGRPHVNSTTSDLSSTFESSMNTPTSI--SSDGMILLSSTTLVTESTTTE 861
 Db 1151 DDPAPVATPTVDK--ITGSTTNGYQVGAAGVGTVEVRDADGTVLGMATTCTDGRKYTVT 1208
 QY 862 LICSDGKECSRLLSSSSGIVTNPDSNESSIVTSTVPT--ASTWSDLSLSDGTSATSSDNV 919
 Db 1209 L--EPGK-ASANETITVAKNATGKESQPATATTPVDLATPTIDTSITGN----- 1254
 QY 920 SKSGVSVT--TETSVTTIQTTPNPLSSSVTSLTQLSSIPSSESESKVTFTSNGDNQSGTH 978
 Db 1255 SKGYEITGTAEPKTTIDVR-----DADGTI 1280
 QY 979 DSQSTSTEIEVITTSSTKVLPPVW-----SSNTDLTSEPTNTRQPTTLSTTSNSITEDIT 1034
 Db 1281 IAATTANE---TGQYTVTLPGAVTTPGETIITISKDGAGNESQPATAVIPADVLAAPT 1336
 QY 1035 TSQPTGDNQNTSSNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGNSGLTTST 1094
 Db 1337 ITKVEGNKANGYTVTGTADPNVTVOFYNSQELLASGNTTGTGTF---SVHIAAGLATEK 1393
 QY 1095 EIEATTSPTE---APSPAVSSGTDVTEPTDTRQPTTLST--TSKTNSELVATV---- 1145
 Db 1394 ETLTALTDTQGNVSPKTTFTMPADITGEPEIKIAAPTSSVLGTSKAGYLKNGTAEPNR 1453
 QY 1146 -----QATNENGKSPSTDLSLTTTGTSTASTSA 1174
 Db 1454 IIQISNRLRSVIAVGATDAEGNFAIQLTAGQATQAQSLATATDGAHYSTATPTMTPA 1513
 QY 1175 NSELVTS--VTGGAVASANDQSHSTSVTNSNSIVSNPQTTLTSSQVTSSTPSTWFTA 1233
 Db 1514 DPTNPGGNGTGGNNGTGGTNGNGATGGNNGSN-----TGSNPNGGGSLG 1563
 QY 1234 STDGSGSI 1242
 Db 1564 TTGSGGLSL 1572

SEQUENCE CHARACTERISTICS:

LENGTH: 1638 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-071-035-262

Query Match 6.0%; Score 392; DB 4; Length 1638;
Best Local Similarity 22.4%; Pred. No. 1.9e-14;
Matches 311; Conservative 149; Mismatches 603; Indels 326; Gaps 60;

QY 13 SIASAKTITGVDFSPNSUTWNAANYAFKPGYPTWNAVL-----GWSLDGTSANPG-DT 66
DB 351 SEVTAITITGV-----TGNSTAGYEVKGT--DANATVEIRNAGGVIGTGTADGTA 401
QY 67 FTLNMPK-VFKYTTTSQTSVDLTADGVKY--ATCFYSGEETFTSTLR-----CFVNDAL 118
DB 402 FTVTVPAGEAGANETLTA VAKNASCETPTTFTQTPADEATVTAPTITGVTGNSAGYEV 461
QY 119 KSSIKAFGTVTLPIAFNVCGTGSSTDLSDSKCFAGTNTVTFNDGDKDISDVEPEKSTV 178
DB 462 KGTADANATVEIR---NAGGVIGTGTADG---TGAFVTVPAGEAGAN-----ETLTA 509
QY 179 DPSAYLYASRYMPSLNKVTTLFVAPQCENGYSCTGMFSSNGDVDAICSNHIGITKGL 238
DB 510 -----VAKNAGSTESTPTTFTQTPADEATVTAPTITGVTGNSAGYEV-----VKGT 554
QY 239 NDWNPVS--SESFSYTKTCSNGIQIKYQNPVAGYRPFIDAYISATDVNQYTLAYTNDY 296
DB 555 ADANATVEIRNAGGAVIGTGTADGTAFTVTAIPAG-----EAGANETLTA VAKN- 603
QY 297 TCAGSRLOSKPFTLAWTGYKNSDAGSNGIVIVATRTVTDSTVATVTLTPENPSVDKTKTI 356
DB 604 ---ASGTESTPTTFQ-----TPADNPVATPIVEITVGTSTTKGYEVKGTAEVG---TTI 652
QY 357 EILQPIPTTTITTSYGVVTSYLTAKTAPIGETATVIVDPYKHTTTVTSEWTTGTTTTT 416
DB 653 EV-RDAAGTVLTATGTGDKY-TVTLDSG-TATA-----NOTLSVAKNAGSTESQAP 704
QY 417 RTNPTD-----SIDTV-----VVQVPLNPVTSTTEYSQSFAFTT-----TVTA 456
DB 705 ATTPADVTAPTVDNITGNSGSGYEITGTADPNTTIEVRDPGSAVIGTGTSDANGDFVTVL 764
QY 457 P-----PGGTDTVIREPNNHTVTTTEYSQSFAFTT-----TVTAPP 494
DB 765 PTGTTNPGDTLVICKDNAGNESQTEVLVPADATVTAPTVGTGNSVAGVQVGTADP 824
QY 495 GGTDSVIREPP-NPTVTTTEYSQSFAFTTTV-TAPPGTDSVIREPNNPTVTTTEYW 552
DB 825 NAT--IEIRDADGNVIATGTADGTGSAFVNLPAAGTANANETLTA LAKDPAGNTSPPTTFQ 882
QY 553 SQSYATTTVTA PGCTDSVIREPNNHTVTTTEYSQSFAFTTTVTAPPG---GTDVVI 609
DB 883 TPA-----DEVAPP-SVDKVTGNTTQGYQVGT-----AELGTTIEVRATDGVLTGAT-- 931
QY 610 IREPNNHTVTTTEYSQSFAFTTTVTPPSTGTDVII-----REPNNPTVTTTEYS 661
DB 932 -----TGPTGQY-----TVTLASGKATQKQTVNVAKNDTGLESQPTTAMTPAD--- 975
QY 662 QSYATTTTI-----TAPPGTDTVIREPNNHTVTTTEYSQSFAFTT--T 704
DB 976 ---VTPPTIGDTGTTGTEYITGTADP--NTTIEVRNPDGTIIGTTTTDDGNTVDLP 1030
QY 705 TVTAPPGETDVLIREPNNHTVTTTEYSQSFAFTT--TVTAPPGETDT---VIIREPPN 759
DB 1031 AGAANPGDTLVVKGDDGNSQPTTEVTPEDATVAAPTIVTGTATGYQVGTGTAEPN 1090
QY 760 PTVTTTEYSQSFAFTT-----TVTAPPGETDVLIREPNNHTVTTTEYSQSFAFTT----- 804
DB 1091 VTIEIHNAGLVAGTGTGAGAFITLTGTATANEALTAIAKDAAGKESNPTAFKTPA 1150
QY 805 --DITSIIPSFSRPHVNVSTTSDLTSTFESSMNTPTSI--SSDGMLLSSTTLVTESETTTE 861

RESULT 14

US-09-071-035-266
; Sequence 266, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 266:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1638 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-071-035-266

Query Match 6.0%; Score 392; DB 4; Length 1638;

Best Local Similarity 22.4%; Pred. No. 1.9e-14;

Matches 311; Conservative 149; Mismatches 603; Indels 326; Gaps 60;

QY 13 SIASAKTICGVDFSNLSWNAANYAFKPGYPTWNAV-----CWSLDGTSANG-DT 66
Db SEVATPTITGV-----TCNSTAGYEVKGTA--DANATVEIRNAGGTIGTGADGTGA 401
QY 67 FTLMNFC-VFKYTTTSOTSDVLTADGVKY--ATCOFYSGEEFTFTSLT-----CTVNDAL 118
Db FTVTVPAGEGANEITLAVAKNASGTEXTPTTFQTPADEATVAPTITGVGTGNSTAGYEV 461
QY 119 KSSIKAFGVTLPIAFNVGSGTSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTV 178
Db KGTADANATVEIR--NAGGTVIGTGTADG---TGAFTVTVPAGEAGAN-----ETLTA 509
QY 179 DPSAYLYASRVPMSLKVITLVAPOCENGYSCTMGFSSNGDVAIDCSNTHIGITKGL 238
Db VAKNASGTESTPTTFQTPADEATVAPTITGVGTGNSTAGY-----VKGT 554
QY 239 NDNNYPVS--SEFSYTKTCTSNIGIQIKYQNPAGYRPFIDAYISATDVNOYTLAYTNDY 296
Db ADANATVEIRNAGGAVIGTADGTGAFVTPAG-----EAGANEITLAVAKN- 603
QY 297 TCAGSLQSKPFLRWYKNSDAGSIVIVATRTVDTSTAVTTLPLFPNSVDKTKTI 356
Db ASGTESPTTFQ-----TPADNPTVATPIVETVGTSTTKGYEVKGTAEVG--TTI 652
QY 357 EILQIPTTTIITSYGVVTSILTKTAPIGETATVIVDVYHTTTVTTSWTGCTTTTTT 416
Db EV--RDAAGVIGTATGTDGKY--TVTLDSG--TATA-----NOTLSYVAKNASGTESQAP 704
QY 417 RTNPTD-----SIDTV-----VVQVPLPNPTVSTTEYWSQSFATTT-----TVTA 456
Db ATPADVTAPTVDNITNGSGYEIETGADPNTIEVRDPGNAVIGTSDANGDTFVL 764
QY 457 P-----PGGTDVVIIEPPNHVTTTEYWSQSFATTT-----TVTAPP 494
Db PTGTTNPGDTLAVIGKDNAGNESQPTPEVLVPADATVAPTIVGTGNSVAGYQVGTADP 824
QY 495 GGTDSVILIREPP--NPVTTTEYWSQSFATTTV--TAPPGTDSVIREPPNPTVTTTEY 552
Db NAT--TEIRDAGNVATGADGTGSFAVNLPAGTANANETLALAKDPAGNTSTPTTQ 882
QY 553 SQSYATTTVTAPPGGTDSVIREPPNHTVTTTEYWSQSYATTTVTAPPG---GTDVI 609
Db TPA-----DEVAPP--SVDAVGTNGTGYQVGTG---AELGTTIEVRATDGVLTAT-- 931
QY 610 IREPPNHVTTTEYWSQSFATTTVTPGSGDVTVII-----REPPNPTVTTTEYWS 661
Db TCGPTGY-----TVTLASGKATAQTVNVAKNDTGLSQPTTAMTPAD--- 975
QY 662 QSATTTTI-----TAPPGTDTVLIREPPNHTVTTTEYWSQSYATTT--T 704
Db VTTPTIGDITGDSYTGITGADP--NTTIEVRNDGTIIGTTTDDQGNFTVDLP 1030
QY 705 TVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTT--TVTAPPGGTDT---VIREPPN 759
Db AGAANFGDILTVVKDGDGNGESQPTVEVPEDATVAAPVTVTGTATGYQVGTAEPN 1090
QY 760 PVTTEYWSQSFATTT-----TVTAPPGGTDTVVIIEYSSSSSKISTSSN----- 804
Db VTEIHNAGLVATGTTGAGAFITLPTGTATANEALTAIAKDAAGKESNPATKPTA 1150
QY 805 --DITSIIIPSRPHVNSTTSDLSFESSMNTPTSI--SSDGMILLSLTLYTESITTE 861
Db ITGSTINGQVVGAAEVGTTVEVRDADGTVLGMATTTGDKGYT 1208

RESULT 15

US-07-757-022B-84

; Sequence 84, Application US/07757022B

; Patent No. 6433142

; GENERAL INFORMATION:

; APPLICANT: Gesner, Thomas G.

; APPLICANT: Clark, Stephen C.

; APPLICANT: Turner, Katherine

; APPLICANT: Hewick, Rodney M.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 Cambridgepark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/757,022B

; FILING DATE: 19910910

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/643,502

; FILING DATE: 18-JAN-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/546,114

; FILING DATE: 29-JUN-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/457,196

; FILING DATE: 29-DEC-1989

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/390,901

; FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 84:
SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-84

Query Match 5.6%; Score 364.5; DB 4; Length 1022;
Best Local Similarity 23.0%; Pred. No. 3.8e-13;
Matches 236; Conservative 109; Mismatches 148; Indels 235; Gaps 45;

QY 252 YTKCTSNQIQIKYQN---VPAGYRPFIDAVISATDVNQYTLAVTNDYTCAGSRQLQ--- 304
DB 73 YESFCAETAVRDNKNRKKRTPPKPPVVDGAGSLDNGDFKVTTPDTSTTQHNKVSISP 132
QY 305 ----SKPFTLRWGTGKNSDAGSNGIVIVATTIV-TDSTTAVTTLFPNPSVDKTKTIEIL 359
DB 133 KITTAKPINRPSLPNSDTSKETSITVNETVETKETTNNKQTSIDGKEKTTSAKET 192
QY 360 QPIPTTTTTSVGVVTSYLKTAFIGETAT--VIVDVPYHTTTTIVSEWTGTI-----TT 413
DB 193 OSIEXT--SAKDAPTSVLAKPTPKAETTKGPAITTPKEPTTPPKPEASTTPKEPTP 250
QY 414 TTRTNPTDSDIDVVVQVPLNPPTVSTTEYWSQSFATTTVTAP-----PGGTDVVIIE 468
DB 251 TTIKSAPT-----TPKEPAPTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTT---KE 301
QY 469 PNHVTVTTEYWSQSFATTTVTAPPGGTDSDVIREPPNPVTTEYWSQSFATTTVTA 528
DB 302 PAPTPTTKSAPTPKPEAPTPPKKAPATPKPEAPTPKPEPTPTT---PKPEAPTTKEPA 357
QY 529 P-----PGGTDSDVIREPPNPVTTEYWSQSFATTTVTAP-----PGGTDSDVIREPPNPVTTEYWSQ 554
DB 358 PTPKPEAPATPKKAPATPKPEAPTPKPEAPTTKEPSPPTPKPEAPTTTKSAPTTTK 417
QY 555 SYATTTTVTAP---PGGTDSDVIREPPNPVTTEYWSQSFATTTVTAP-----PG 603
DB 418 EPAPTTTKSAPTPKPEAPTPPKKAPATPKPEAPTPKPEAPTTT---PKPEAPTTKEPA 475
QY 604 GDT-----VIREP-----TEYWSQSFATTTVTGPPSGDVTVIIEP 649
DB 476 PTTTKKAPATPKPEAPTPKPEAPTPPKKAPATPKPEAPTPKPEAPTTPEELAPTP 535
QY 650 PNPTVTTEYWSQSFATTTVTAP---PGGTDVLIIEP-----PNHVTVTTEYWSQSFATTT 703
DB 536 EEPPTTPE---EPAPTTPKAAAPNTPKPEAPTPKPEAPTTKPEAPTTKPEAPTTT---KETAPT 590
QY 704 TTVTAP-----PGGTDVLIIEP-----PNHVTVTTEYWSQSFATTT 736
DB 591 PKGTAPTTKPEAPTPPKKAPATPKPEAPTTTKEPTSTTSKAPATPKGTAPTP---KEPA 648
QY 737 ATTTTVTAP---PGGTDVVIIEPPNPVTTEYWSQSFATTTVTAPPGGTDVVIIEP 794
DB 649 PTPKPEAPTPKGTAPTTKPEA-PT-TPKKAPKELAPTTT---KGPTSTSDKAP 702
QY 795 SSSKLSSTSDNITSIIIPFSRPHYVNSTSDLSSTFESSMNPSTSISSDGMLLSSTTLVT 854
DB 703 TTPKETAPT---TPKPEAPTPKPEAPTTTPEPTTSEVSTPT-----T 744
QY 855 ESETTELICSDGKCSRLSSSGIVTNPDSNESSIVTSTVPTASTMSDLSSTDGISAT 914
DB 745 TKEPT-----IHKSPD---ESTPELSAEPKALENS-PKEPGVPTT 783
QY 915 SSDNVSKGVSVTTTSTVT--TIQTPNPLSSSVTSLTQLSIPVSESESKVTTSTNGD 972

DB 784 KTPAATKPEMTTAKDKITTERDLRTPE-TTTAAPKMTKETATTTTEKTTESKITATTTQV 842
QY 973 NOSGTHDSQSTSTEIEIVTTSKVLPPVSS-NTDLTSEPTNTREQPTTLSTTSNSITE 1031
DB 843 TSTTTQD-----TTPFKITTLKTTTLAPKVTITTKTITTEIMNKPEE----- 885
QY 1032 DITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKGSGSHESASTSLKPSMGENSEGLT 1091
DB 886 ---TAKPK-DRATNSKATTPKPKPT---KAPKPTSTKKPKTMRVRKPK-----T 930
QY 1092 TSTEIEATTSPTSPAPVSSG-TDVTTEPTDTRREQPTTLSTTSKTNSELVATTOATNE 1150
DB 931 TPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQT-----PNSKLVNPKSED 978
QY 1151 NGGKSPST 1158
DB 979 AGGAGET 986

Search completed: June 11, 2003, 17:12:35
Job time : 30 secs

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	4078.5	57.8	1047	3	O93865		O93865 candida alb
2	3716.5	52.2	1523	3	O9HFx4		O9HFx4 candida alb
3	2791.5	43.0	1443	3	O9Y8F2		O9Y8F2 candida alb
4	2673.5	41.2	1270	3	O9Y743		O9Y743 candida alb
5	2576.5	39.7	2297	3	O9HGK6		O9HGK6 candida alb
6	1986.5	30.6	468	3	O9URQ0		O9URQ0 candida alb
7	1823	28.1	469	3	O9URP8		O9URP8 candida alb
8	1780.5	27.4	468	3	O9c471		O9c471 candida alb
9	1471	22.6	336	3	O9HF70		O9HF70 candida dub
10	1191	18.3	331	3	O9HF72		O9HF72 candida dub
11	1176	18.1	331	3	O9HF69		O9HF69 candida dub
12	999.5	15.4	433	3	O9Y8F1		O9Y8F1 candida alb
13	981	15.1	353	3	O9HF71		O9HF71 candida tro
14	831.5	12.8	1195	3	O9GW08		O9GW08 schizosacch
15	834	12.7	226	3	O9HG16		O9HG16 candida alb
16	698.5	10.8	1079	5	O9N4S7		O9N4S7 caenorhabdi

```
QY 181 SAYLYASRVMPSLNKVTTLEVAPOCENGYSYSGTMGFSSNGDVADICSNHIGITKGLND 240
Db 181 KGYLTDTSRVIPSLNKVSTLVAPOCANGYSYSGTMGFANTYGDVQIDCSNHHVIGITKGLND 240
QY 241 WNPVSESESYTKTCSNGIQKYNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db 241 WNPVSESESYTKTCSNGIFITYKNVPAGYRPFVDAYISATDVNSYTLISYANEYTCAG 300
QY 301 SRLQSKPFTLRWGTGYNKSDAGSNGIVVATRTVTDSTTAVTTLTPNPSVDKTKTLEILQ 360
Db 301 GYQWRAPFTLRWGTGYNKSDAGSNGIVVATRTVTDSTTAVTTLTPNPNKDKTKTLEILQ 360
QY 361 PIPTTTITTSYGVGTTSYLTKTAPIGETATVIVDVPYHTTTVTSEWGTITTTTATNP 420
Db 361 PIPTTTITTSYGVGTTSYLTKTAPIGETATVIVDVPYHTTTVTSEWGTITTTTATNP 420
QY 421 TDSIDTVVQVPLPNTVSTTEYWSQSFATTTVTAPPGTDTVIREPPNHTVTTTEY 480
Db 421 TDSIDTVVQVPLPNTVSTTEYWSQSFATTTVTAPPGTDTVIREPPNHTVTTTEY 480
QY 481 SQSFATTTVTAPPGTDSVIREPPNHTVTTTEYWSQSFATTTVTAPPGTDSVIRE 540
Db 481 SESYTTTSTTAPPGTDSVIREPPNHTVTTTEYWSQSFATTTVTAPPGTDSVIRE 540
QY 541 PPNTVTTTEYWSQSFATTTVTAPPGTDSVIREPPNHTVTTTEYWSQSFATTTVT 600
Db 541 PPNTVTTTEYWSQSFATTTVTAPPGTDSVIREPPNHTVTTTEYWSQSFATTTVT 600
QY 601 PPGGTDVIREPPNHTVTTTEYWSQSFATTTVTGPPGTDVIREPPNHTVTTTEY 660
Db 601 PPGGTDVIREPPNHTVTTTEYWSQSFATTTVTGPPGTDVIREPPNHTVTTTEY 660
QY 661 SQSYATTTTITAPPGTDTVIREPPNHTVTTTEYWSQSFATTTVTAPPGTDTVIRE 720
Db 625 SQSYATTTTITAPPGTDTVIREPPNHTVTTTEYWSQSFATTTVTAPPGTDTVIRE 684
QY 721 PPNTVTTTEYWSQSFATTTVTAPPGTDTVIREPPNHTVTTTEYWSQSFATTTVT 780
Db 685 PPNTVTTTEYWSQSFATTTVTAPPGTDTVIREPPNHTVTTTEYWSQSFATTTVT 744
QY 781 PPGGTDVIREPPNHTVTTTEYWSQSFATTTVTGPPGTDVIREPPNHTVTTTEY 840
Db 745 PPGGTDVIREPPNHTVTTTEYWSQSFATTTVTGPPGTDVIREPPNHTVTTTEY 776
QY 841 SSDGMLLSSTLTETESETTELICSDGKCSRLSSSGIVTNPDSNESSITVTSTVPTAST 900
Db 777 -----LWSTTWIEKTIETSCGDKGCSWVSSTRIVIPNNIETPNVTVDSSTT 830
QY 901 MSDSLSTGDISATSSDNVSKGVSYTTETS-VTTIQTTPNPLSSSVTSLQLSSIPSVS 959
Db 831 ESTS-QSPSGI-----FSESGVSSETSTVTTAQTN-----PSVP 865
QY 960 ESESKVTFTSNGDNQSGTHDSQSTSEIEIVTTSSTKVLPPVYSSNTDLTSEPTNTREQ 1019
Db 866 TTESEVEFTKGNNGNGPYESPSTH-----YKSSMDENSEPT-----902
QY 1020 TTLSSTNSITDITTSQPTGONGDNTSSTNPVPTVATSTLASAEDNKSGSHESASTS 1079
Db 903 -----TSTAAS-----908
QY 1080 LKPSMGENSELTTSTEIE---ATTSTPEAPSPAVSGGTDVTEPTDTRQPTLTSTTK 1136
Db 909 -----TSDIENATIAATGVSSEASSPIIISADET-----TTITTAET 946
QY 1137 TNSLVATTOATNEN-CKGKSPSDDLTSLLTGTSTASANSSELVTSVSGVTGGAVASAND 1195
Db 947 STSVI---EOPTNNGGKAPSA-----TSSPSTTTTANNDVSITG-----TSTN 989
QY 1196 QSHSTVSTNSNSIVSNTPOTLLSQOVTSSSPSTNTFIASDTGSGSIIQHSWLYGLITL 1255
Db 990 QSQSQSQSN-----SDTQOQTLTSSQMTSSLSVLSH--MLTTFDGSVSGVIOHSWLCGLITL 1042
QY 1256 LSLFI 1260
```

```
Db 1043 LSLFI 1047
RESULT 2
Q9HFX4
ID Q9HFX4 PRELIMINARY; PRT; 1523 AA.
AC Q9HFX4;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE Agglutinin-like protein (Fragment).
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC5314;
RA Chen X., Chen J.-Y.;
RT "ALS4 (agglutinin-like sequence) of Candida albicans.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF272027; AAG25054.1; -
FT NON_TER 1523
SQ SEQUENCE 1523 AA; 159168 MW; 73AF3B3E442FD53C CRC64;
Query Match 57.2%; Score 3716.5; DB 3; Length 1523;
Best Local Similarity 52.0%; Pred. No. 5.7e-148;
Matches 798; Conservative 148; Mismatches 300; Indels 289; Gaps 30;
QY 1 MLQOFTLLFLYLSIASAKTITGVDFSNLTSWNAAYAPKPGYPTWNAVLGWSLDGTS 60
Db 1 MLQOFTLLFLYLSIASAKTITGVDFSNLTSWNAAYAPKPGYPTWNAVLGWSLDGTS 60
QY 61 ANPGDFTLNMPCVKFYKTSQTSVLDADGKVATCOFYSGEEFTFTSLTCTVNDALKS 120
Db 61 ASAGDFTLDMPCVFKFITDQTSIDLVAQDRTVATCNLSAEFTTFFSSVCTVTTMTA 120
QY 121 SIKAFGTVTLPIAFNVGGTGSSTDLSDSKCFTAGTNTVTFNDGDKDISIDVEKSTVDP 180
Db 121 DTAKIAGTVTLPIAFNVGGTGSSTDLSDSKCFTAGTNTVTFNDGDKDISIDVEKSTVDP 180
QY 181 SAYLYASRVMPSLNKVTTLEVAPOCENGYSYSGTMGFSSNGDVADICSNHIGITKGLND 240
Db 181 SORILLRILPILSOAVSLFLPOECANGYSGTMGFSTAGTGATIDCSVHVGISGLND 240
QY 241 WNPVSESESYTKTCSNGIQKYNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db 241 WNPVSESESYTKTCSNGIQKYNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
QY 301 SRLQSKPFTLRWGTGYNKSDAGSNGIVVATRTVTDSTTAVTTLTPNPSVDKTKTLEILQ 360
Db 301 ASVDSDFTWWSGYSNQSAGSNGIIVVTRTVTDSTTAVTTLTPNPSVDKTKTLEILQ 360
QY 361 PIPTTTITTSYGVGTTSYLTKTAPIGETATVIVDVPYHTTTVTSEWGTITTTTATNP 420
Db 361 PIPTTTITTSYGVGTTSYLTKTAPIGETATVIVDVPYHTTTVTSEWGTITTTTATNP 420
QY 421 TDSIDTVVQVPLPNTVSTTEYWSQSFATTTVTAPPGTDTVIREPPNHTVTTTEY 480
Db 421 TDSIDTVVQVPLPNTVSTTEYWSQSFATTTVTAPPGTDTVIREPPNHTVTTTEY 480
QY 481 SQSFATTTVTAPPGTDSVIREPPNHTVTTTEYWSQSFATTTVTAPPGTDSVIRE 540
Db 481 SQSFATTTVTAPPGTDSVIREPPNHTVTTTEYWSQSFATTTVTAPPGTDSVIRE 540
QY 541 PPNTVTTTEYWSQSFATTTVTAPPGTDSVIREPPNHTVTTTEYWSQSFATTTVT 600
Db 541 PPNTVTTTEYWSQSFATTTVTAPPGTDSVIREPPNHTVTTTEYWSQSFATTTVT 600
QY 601 PPGGTDVIREPPNHTVTTTEYWSQSFATTTVTGPPGTDVIREPPNHTVTTTEY 660
Db 601 PPGGTDVIREPPNHTVTTTEYWSQSFATTTVTGPPGTDVIREPPNHTVTTTEY 660
```


[illegible]


```
QY 1027 NSITEDITTSQPTGONG-----DNTSSNPVPTVATSTLASASEEDNKGSHESASTSL 1080
Db 997 SSHSEIFSS-----DNSVLKQVDRESTITKSTPTDVTTSVLSVHSTE-----ASTA- 1044
QY 1081 KPSMGENS-----GLTSTSEIATTSPTAPSPAVSSGT-----DVTTEPTDT- 1124
Db 1045 --TLGENSFNSVASTPLNATSLRSTSSSNHATE---SSGTVKSEASVEALPSPPTSTD 1099
QY 1125 -----REQPTTLSTSTKTN-----SELVATTOATN---ENGGKSPSTDLTSSLTGCT 1168
Db 1100 NPLSYSTEAEGITVANGSTNNLITESQVAAPTDSTSVLIENLVVSTFDDNSSAAVDQ 1159
QY 1169 SASTSANGSELV---TSGSVTGGAVASANDQSHSTSV-----TNSNSI 1208
Db 1160 PSKTKSIEESINPDSNETNGFIATLSQAQVPSSIHSELISLITTAKTDDASNMGDSA 1219
QY 1209 VSNTPQTLTQQVTSSTSTNFIATSTVDGSGSIIOHSTWLTGLTLISLF 1259
Db 1220 ASNSQPTTLTQQVATS--SYNOPLITTVAGSSSATKHPSWLLKFISVALFF 1268

RESULT 5
Q9HGK6
ID Q9HGK6 PRELIMINARY; PRT: 2297 AA.
AC Q9HGK6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Agglutinin-like protein Als7p.
GN ALS7.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RX MEDLINE=20321177; PubMed=10861907;
RA Hoyer L.L., Hecht J.E.;
RT "The ALS6 and ALS7 genes of Candida albicans.";
RL Yeast 16:847-855(2000).
DR EMBL; AF201684; AAF98068.1; -.
SQ SEQUENCE 2297 AA; 244723 MW; 59B020C63027F651 CRC64;

Query Match 39.7%; Score 2576.5; DB 3; Length 2297;
Best Local Similarity 41.3%; Pred. No. 4.3e-100;
Matches 595; Conservative 220; Mismatches 416; Indels 211; Gaps 24;

QY 2 LQOFTLLFLYLSIAS--AKTITGVDFSNLSLFWNSAANYAFKPGPYPTWNAVLGWSLDGT 59
Db 1 MKKLLVLLASFTTVISKEVTGVFNQFNLSWSTYRAREYEEISLTLANAQLWALDGT 60
QY 60 SANPGDTTLNMPGVKVTTSQTSVDLTADGVKYATCQPYSGEFTFTSLTCTVNDALK 119
Db 61 IASPGDTTLNMPGVKVTYTSVQLTANSTAYATCDFDAGEDTKSFSLKCTVDELT 120
QY 120 SSIAFGVTLPIAFNVGVTGSGTLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVD 179
Db 121 EDTSVFGSVILPIAFNVGSGSKSITDSKCFSSGYNVTFDFGNNQLSTTANFLPRRL 180
QY 180 PSAYLYAGRVMSPLNKVTLTFVAPQCENGYTSGTMGSSNGDVAIDCSNIHIGITKGLN 239
Db 181 AFLVLVSQRLMSLDTMNFVMSVPCFMCYQSGKLGFTSNDIDDFEIDCSIIHVGINEIN 240
QY 240 DWNYPVSSSFYTKTCTNSGIIQKYQNPAGYRPFIDAYISATDVNQVTLAYTNDYCA 299
Db 241 DWSMVSVPFDHTIRCTSRALYIEFKTIPAGYRPFDAIVQIPTTEPFVRYTNEFACV 300
QY 300 GSRLOSKPFTLRWT--GYKNSDAGSGIVVATRVVTDSTTAVTTLPNPNSVDKTKTIEI 358
Db 301 NGIYTSIPFTSFFSQPILYDEALAIAGDLVRTSTVIGSITRTTLTLPFISLQKTKTILV 360
QY 359 LQPIPTTTITTSYGVGVTTSYLTKTAPIGETATVIVDPYHTTTTTSWTGTITTTTTRT 418
```

```
Db 361 LEPIPTTTTTHHGFDTWYTKKATIGDTATVFDVQPORTATTLTYQESSTATTTVF 420
QY 419 NPTDSLDTVVQVPLP-----NPTVSTTE 442
Db 421 DDIDLVDVTVIVKIPYNPNTIITQFWSGKYLTETHEKPEPLGSDSVIIRKPHNPTVTTE 480
QY 443 YMSQSFATTTVTAPPGGTDVLIIRPPNHTVTTEYWSQSFATTTTVPAPGCGTDSVII 502
Db 481 FWSEFATTTTINYPEGSDSVIIRPPNHTVTTEYWSQSFATTTTINYPEGCGTDSVIV 540
QY 503 REPPNPTVTTEYWSQSFATTTTVPAPGCGTDSVIIIRPPNHTVTTEYWSQSFATTTT 562
Db 541 REPHNPTVTTEFWSEFATTTITNYPEGSDSVIIRPPNHTVTTEFWSEFATTTETI 600
QY 563 TAPPGCGTDSVIIIRPPNHTVTTEYWSQSFATTTTVPAPGCGTDSVIIIRPPNHTVTTE 622
Db 601 TNGPEGSDSVIIRPPNHTVTTEFWSEFATTTITNYPEGCGTDSVIIIRPPNHTVTTK 660
QY 623 YMSQSFATTTVTGPPGSDTVLIIRPPNHTVTTEYWSQSFATTTTITAPPGTDTVLI 682
Db 661 FWSEFATTTTINYPEGSDSVIIRPPNHTVTTEFWSEFATTTITNYPEGCGTDSVIV 720
QY 683 REPPNHTVTTEYWSQSFATTTTVPAPGCGTDSVIIIRPPNHTVTTEYWSQSFATTTT 742
Db 721 REPHNPTVTTEFWSEFATTTITNYPEGSDSVIIRPPNHTVTTEFWSEFATTTETV 780
QY 743 TAPPGCGTDSVIIIRPPNHTVTTEYWSQSFATTTTVPAPGCGTDSVIIIRPPNHTVTTE 798
Db 781 TNYPEGSDSVIIRPPNHTVTTEFWSEFATTTITNYPEGCGTDSVIIIRPPNHTVTTK 840
QY 799 ISTSSNDITSI-----IPS-----FSRP 816
Db 841 IESSDGNISSAQESSSSVEQFSADETSSIVELSSRDIPSSSGLTSSSSTVSSYD 900
QY 817 HYVNSTT-----SDLSTFESSMNTPTSSD--GMLLSSTT----- 851
Db 901 SYSSSTSESSIASVSDYSSESSIESSTLSSSDRYSSSISDITTSFWDSSSLESTITS 960
QY 852 -----LVTESETTELICSDGKECRLSSSGIIVNPDNSESIVTSTVPTAS 899
Db 961 SSIDAQSHLVQVSNISITQFISSSSEESTSATDALVSDASSILSDTSYSPSS 1020
QY 900 TMSDLSLSTDGISATSSDNVSKSGSVTTTSTVTTIQTTPNPLSSSVTSLTQ-----SSI 955
Db 1021 TISPDDEFHTI-AGESDSQSISFITSTVEISDSVSLTSDP-ESSFDSRSLNSDSSSS 1078
QY 956 PS-----VSESESKVTFTSNGDNQSGTHDSQSSTSTEIEIVTTSSTKVLPP----- 1000
Db 1079 PSTDQDRDLTSSFSFLIKSGSGRESSIGTILSESSSDSIPTTFTSTRYWSPSGMSRHVT 1138
QY 1001 -----VSSNT--DLTSEPTNTREOPTTLSTTSNSITEDITTSQPTGONGDWTSS 1049
Db 1139 NSTETSDVSVSSVAGDETSSESVSVISESESVTSESVASESVASESVASESVASESV 1198
QY 1050 NPVPVATSLASAEEDNKGSHESASTSLKPSMGENGSLTSTSEIATTSPTAPSP 1109
Db 1199 TAVSDI--SDLYTTEVSTSDNSGMSPI-PSSEQRS-----SIPIMSSDESESR 1249
QY 1110 AVSSGTDVTEPTDTRREQPTTLST-----TSKTNSELVATTOATNENGG 1153
Db 1250 ESSSGTILSEENSDS--IPTFTSTRYWSPSGMSRHVTSTSTSVSDVSVSSVAGDETSE 1307
QY 1154 KPSF--TDLTSLTCTGTSASTSANSSELVTSVSGVTGCAVASANDQSHSTSVTNSN- 1207
Db 1308 SSVSVISESESVTSESVASESVASESVASESVASESVASESVASESVASESVASESV 1367
QY 1208 -----IVSNTPQ-----TTLSQOVTSSSPSTNTIASTYDGGSGIIIOH 1245
Db 1368 PIPSSBQRSIPVMSSDESSSESRESSESSTILSEENSISPT--TFTSTRYLSPSGMSRRH 1425
QY 1246 ST 1247
```


DR EMBL; AF229989; AAK00764.1; -

FT NON_TER 468

SQ SEQUENCE 468 AA; 50127 MW; B291D3EB15FB96DE CRC64;

Query Match 27.4%; Score 1780.5; DB 3; Length 468;
Best Local Similarity 70.1%; Pred. No. 1.7e-67;
Matches 329; Conservative 62; Mismatches 77; Indels 1; Gaps 1;

QY 1 MLQFTLLFLYLSTASAKTITGVDFSNLSNAANYAFKPGVPTWNAVGLWGLDGTG 60

DB 1 MLQFILLFISLTSTAKITGVFNFSFSLTWTSRVAYKGPETPTWNAVGLWGLSNTT 60

QY 61 ANPGDTFTLNPMPCKYKTTSTQTSVDLTADGVKYATCFQYSGEEFTTFLTCTVNDALKS 120

DB 61 ADPGDTFTLLPCVFKFTTQTSVDLTADGVSYATCDFNAGEEFTTFSLSCTVNSVSV 120

QY 121 SIKAFGTVTLPIAFNVGGTGSTDLSDSKCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180

DB 121 YARVSGTVKLPITFNVGGTGSVDLADSKCTAGKNTVTFMDGDKTISTTVDFDASPVP 180

QY 181 SAYLYASRVMPSLNKTTLFVAPQCENGYTSCTMGFSSNGDVAIDCSNIHIGITKGLND 240

DB 181 SGYITSSRIIFSLNKLSSLFVVPQCENGYTSCTMGFVSNAG-ATIDCSNVNIGISKGLND 239

QY 241 WNPVSSSEFSYTKTCTSNGLIOIKYQNPAGYRPFIDAYISATDYNQYTLAYTNDYTCAG 300

DB 240 WNPVSSSEFSYTKTCTSTVTFQNPAGYRPFVDAISAENIDKYTLTYANEYTCEN 299

QY 301 SRLQSKPPTLRWTKYKNSDAGSNGIVIVATTTVDSTAVTTLFPNPSVDKTKTIEILQ 360

DB 300 GNTVVDPTLLTMWYKNSDAGSNGIVIVATTTVDSTAVTTLFPNPSVDKTKTIEILQ 359

QY 361 PIPTTTITTSYGVVTTSLTKTAPIGETATVIVDPYHTTTTSEWGTITTTTTRNP 420

DB 360 PIPTTTITTSYGVVTTSLTKTAPIGETATVIVDPYHTTTTSEWGTITTTTTRNP 419

QY 421 TDSIDTVVQVPLPNPTVTTTEYMSOSFATTTTVAPEGGTDVIREP 469

DB 420 TGSIDTVIVQIPSPDPTTTEFWSESPASTTTINPPDGTNSVIREP 468

RESULT 9

Q9HF70

ID Q9HF70 PRELIMINARY; PRT; 336 AA.

AC Q9HF70;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE Agglutinin-like protein Als2p (Fragment).

GN ALS2.

OS Candida dubliniensis (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; mitosporic Saccharomycetales; Candida.

OX NCBI_TaxID=42374;

RN SEQUENCE FROM N.A.

RC STRAIN=CD36;

RA Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;

RT "Evidence suggesting the presence of an ALS gene family in Candida

RT dubliniensis and Candida tropicalis.";

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF202529; AAG35623.2; -

FT NON_TER 1

FT NON_TER 336

SQ SEQUENCE 336 AA; 35593 MW; E513E6EA9E89EC7 CRC64;

Query Match 22.6%; Score 1471; DB 3; Length 336;

Best Local Similarity 81.0%; Pred. No. 1.1e-54;

Matches 272; Conservative 31; Mismatches 33; Indels 0; Gaps 0;

QY 61 ANPGDTFTLNPMPCKYKTTSTQTSVDLTADGVKYATCFQYSGEEFTTFLTCTVNDALKS 120

DB 1 ASPGDTFTLNPMPCKYKTTSTQTSVDLTADGVKYATCFQYSGEEFTTFLTCTVNSALTS 60


```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1177;
RA Hoyer L.L.; Kapteyn J.C.; Hecht J.E.; En J.; Klis F.M.;
RT "Evidence Suggesting the Presence of an ALS Gene Family in Candida
RL dubliniensis and Candida tropicalis.";
DR EMBL; AFI89016; AAG17111.1; -
FT NON_TER 1
FT NON_TER 226
SQ SEQUENCE 226 AA; 24802 MW; 2AA36A42E44CB460 CRC64;

Query Match 12.7%; Score 824; DB 3; Length 226;
Best Local Similarity 70.1%; Pred. No. 8.3e-28;
Matches 155; Conservative 28; Mismatches 38; Indels 0; Gaps 0;

-QY 429 VQVPLNPVSTTEYWSQSFATTTVTAPPGGTDVVIIREPPNHVTTTEYWSQSFATTT 488
Db 1 VQVPSNPVTVTTFQWSSGVPVTTVTGPGTDSVVIKEPHNPVTTFEWSQSFATTE 60
-QY 489 TVTAPPGGTDVVIIREPPNPVTVTTEYWSQSFATTTVTAPPGGTDSVVIIREPPNPVT 548
Db 61 TVTNNEGDSVVIKEPHNPVTTFEWSQSFATTTVTNYPEGTDSVIVREPHNPVT 120
-QY 549 TEYWSQSYATTTVTAPPGGTDSVVIIREPPNHVTTTEYWSQSYATTTVTAPPGGTDTV 608
Db 121 TEWSESFATTTVTNYPEGTDSVIVREPHNPVTTFEWSQSFATTTITNYPEGTDSV 180
-QY 609 IIREPPNHVTTTEYWSQSFATTTVTGPGSGTDVVIIREP 649
Db 181 IVREPHNPVTTFEWSQSFATTTITGTLGTDVSIHDP 221
```

Search completed: June 11, 2003, 17:15:00
Job time : 61 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:11:32 ; Search time 29 Seconds
(without alignments)
4176.875 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQOFTLLFLYLSIASAKTI.....SIQHSTWLYGLTLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	6495	100.0	1260	2 S60896	agglutinin-like pr
2	3194	49.2	1419	2 T30531	agglutinin-like ad
3	939	14.5	1367	2 S51959	hypothetical prote
4	939	14.5	1537	2 S53465	flocculation prote
5	711	10.9	1075	2 S48992	flocculation prote
6	675.5	10.4	1367	1 S48478	glucan 1,4-alpha-g
7	649.5	10.0	1275	2 T33369	hypothetical prote
8	646.5	10.0	2232	2 T34434	hypothetical prote
9	637.5	9.8	2271	2 F90073	hypothetical prote
10	632.5	9.7	3570	2 T45025	mucin MUC5B, trach
11	599	9.2	3020	2 A43932	mucin 2 precursor,
12	578	8.9	1283	2 T39174	hypothetical Serin
13	577	8.9	1161	2 S57180	probable membrane
14	553.5	8.5	1459	2 T32271	hypothetical prote
15	547.5	8.4	1609	2 S25345	probable membrane
16	526	8.1	1229	2 T25697	hypothetical prote
17	526	8.1	1630	2 A53577	ascites stialoglyco
18	515	7.9	4776	2 E95206	cell wall surface
19	511.5	7.9	3507	2 T34513	hypothetical prote
20	503.5	7.8	1777	2 T34369	hypothetical prote
21	501	7.7	1169	2 S38181	flocculation prote
22	492.5	7.6	1104	2 S59310	probable membrane
23	487	7.5	1832	2 T31113	mucin-like glycopr
24	482.5	7.4	1829	2 T24583	hypothetical prote
25	464.5	7.2	670	2 F36791	hypothetical prote
26	460	7.1	770	2 T22808	hypothetical prote
27	451.5	7.0	786	2 T16509	hypothetical prote
28	445	6.9	1802	2 S69703	HKRI protein precu
29	443.5	6.8	1032	2 T34433	hypothetical prote

30	443.5	6.8	1251	2 T21389	hypothetical prote
31	435.5	6.7	2468	2 A83412	hypothetical prote
32	430.5	6.6	866	2 T45462	membrane glycoprot
33	428	6.6	867	2 T45463	membrane glycoprot
34	424	6.5	528	2 T47141	gastric mucin (clo
35	423.5	6.5	520	2 S62521	hypothetical prote
36	419	6.5	973	2 T40778	hypothetical 129.5
37	418.5	6.4	725	2 A41258	a-agglutinin core
38	415.5	6.4	1306	2 S25370	MSB2 protein - yea
39	412.5	6.4	3029	2 S78109	hypothetical prote
40	412	6.3	851	2 T22696	hypothetical prote
41	409.5	6.3	1041	2 S55862	probable membrane
42	405	6.2	833	2 E90577	lipoprotein vsac l
43	404.5	6.2	2035	2 A40718	host cell factor C
44	401.5	6.2	648	2 PC4395	mucin 3 - human (f
45	400.5	6.2	948	2 T11678	hypothetical prote

ALIGNMENTS

RESULT 1

S60896
agglutinin-like protein - yeast (Candida albicans)
C:Species: Candida albicans
C:Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-2000
C:Accession: S60896
R:Hoyer, L.L.; Scherer, S.; Shatzman, A.R.; Livli, G.P.
Mol. Microbiol. 15, 39-54, 1995
A:Title: Candida albicans ALS1: domains related to a Saccharomyces cerevisiae sexual
A:Reference number: S60896; MUID:95272392; PMID:7752895
A:Accession: S60896
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1260 <HO>
A:Cross-references: EMBL:L25902; NID:g704426; PIDN:AAC41649.1; PID:g704427
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosid

Query Match 100.0%; Score 6495; DB 2; Length 1260;
Best Local Similarity 100.0%; Pred. No. 1.7e-273; Indels 0; Gaps 0;
Matches 1260; Conservative .0; Mismatches 0;

QY	1	MLQOFTLLFLYLSIASAKTITGVDFSNLSITWSNAANYAFKGPYPTWNAVLGWSLDGTS	60
DB	1	MLQOFTLLFLYLSIASAKTITGVDFSNLSITWSNAANYAFKGPYPTWNAVLGWSLDGTS	60
QY	61	ANPGDTFTLNMPGVFKYTTTSQTSVDLTADGVKYATCOFYSGEFTTFTSLTCTVNDALKS	120
DB	61	ANPGDTFTLNMPGVFKYTTTSQTSVDLTADGVKYATCOFYSGEFTTFTSLTCTVNDALKS	120
QY	121	SIKAFCTVTLPTAFNVGGTSSSTDLEDKSCFTAGTNTVTNFDGDKDISIDVEFEKSTVDP	180
DB	121	SIKAFCTVTLPTAFNVGGTSSSTDLEDKSCFTAGTNTVTNFDGDKDISIDVEFEKSTVDP	180
QY	181	SAYLYASRVMPSLNKVTTTLFVAPQCENGYSSTGMFSSSSNGDVAIDCSNIHIGITKGLND	240
DB	181	SAYLYASRVMPSLNKVTTTLFVAPQCENGYSSTGMFSSSSNGDVAIDCSNIHIGITKGLND	240
QY	241	WNYPVSSSEFSYTKTCTSNIGIQIKYONVPAGRPFFIDAYISATDVNOYTLAYTNDYTCAG	300
DB	241	WNYPVSSSEFSYTKTCTSNIGIQIKYONVPAGRPFFIDAYISATDVNOYTLAYTNDYTCAG	300
QY	301	SRLOSKPFTLRWTGYKNSDAGSNGIVIVATTRVTDSTTAVTTLTPNPSVDKTKTTEILQ	360
DB	301	SRLOSKPFTLRWTGYKNSDAGSNGIVIVATTRVTDSTTAVTTLTPNPSVDKTKTTEILQ	360
QY	361	PIPTTTITTSYGVVTSYLTAKPIGETATVIVDVPYHTTITVTSWTGTTITTTTTRNP	420
DB	361	PIPTTTITTSYGVVTSYLTAKPIGETATVIVDVPYHTTITVTSWTGTTITTTTTRNP	420
QY	421	TDSIDTVVQVPLPNPTVSTTEYSQSFAITTTVTAPPDGTDTVIIREPNNHTVTTEYW	480
DB	421	TDSIDTVVQVPLPNPTVSTTEYSQSFAITTTVTAPPDGTDTVIIREPNNHTVTTEYW	480

```
QY 481 SQSFATTTTAPPGTDSVIREPPNPVTTTTEYWSQSFATTTTVPAPGGTDSVIRE 540
|||||
Db 481 SQSFATTTTAPPGTDSVIREPPNPVTTTTEYWSQSFATTTTVPAPGGTDSVIRE 540
QY 541 PNPVTTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 600
|||||
Db 541 PNPVTTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 600
QY 601 PPGTDSVIREPPNPVTTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEY 660
|||||
Db 601 PPGTDSVIREPPNPVTTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEY 660
QY 661 SQSFATTTTAPPGTDSVIREPPNPVTTTTEYWSQSFATTTTVPAPGGTDSVIRE 720
|||||
Db 661 SQSFATTTTAPPGTDSVIREPPNPVTTTTEYWSQSFATTTTVPAPGGTDSVIRE 720
QY 721 PNPVTTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 780
|||||
Db 721 PNPVTTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 780
QY 781 PPGTDSVIREPPNPVTTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEY 840
|||||
Db 781 PPGTDSVIREPPNPVTTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEY 840
QY 841 SSGMLSSSTLVTESFETTELICSDGKESRSLSSSGIVTNPDSNESSIVTSTVPTAST 900
|||||
Db 841 SSGMLSSSTLVTESFETTELICSDGKESRSLSSSGIVTNPDSNESSIVTSTVPTAST 900
QY 901 MSDSLSTDGISATSDNVSQSGSVTSTETSTETIIVTTSKVLPPVSSNTDLTSEPTNTREOPT 960
|||||
Db 901 MSDSLSTDGISATSDNVSQSGSVTSTETSTETIIVTTSKVLPPVSSNTDLTSEPTNTREOPT 960
QY 961 SESKVFTSNGDNQSGTHDSQSTSTETIIVTTSKVLPPVSSNTDLTSEPTNTREOPT 1020
|||||
Db 961 SESKVFTSNGDNQSGTHDSQSTSTETIIVTTSKVLPPVSSNTDLTSEPTNTREOPT 1020
QY 1021 TLSTTNSITPEDITTOPTGDNDGNTSSNPVPTVATSTLASASEEDNKGSGHESASTSL 1080
|||||
Db 1021 TLSTTNSITPEDITTOPTGDNDGNTSSNPVPTVATSTLASASEEDNKGSGHESASTSL 1080
QY 1081 KPSMGESGLTTSTETIATTTSTAPSPAVSGTDTTPTDTRQPTTLSTSTKNSE 1140
|||||
Db 1081 KPSMGESGLTTSTETIATTTSTAPSPAVSGTDTTPTDTRQPTTLSTSTKNSE 1140
QY 1141 LVATTQATNENGKSPSTDLTSLTCTASTSANSSELVTSVGTGAVASANDQSHST 1200
|||||
Db 1141 LVATTQATNENGKSPSTDLTSLTCTASTSANSSELVTSVGTGAVASANDQSHST 1200
QY 1201 SVTNSNSIVNTPOTTLSSQVTSSTNTFFIASTYDGGSGIIQHSWTWLYGLITLISLFI 1260
|||||
Db 1201 SVTNSNSIVNTPOTTLSSQVTSSTNTFFIASTYDGGSGIIQHSWTWLYGLITLISLFI 1260
```

RESULT 2

T30531

agglutinin-like adhesin - yeast (Candida albicans)

C:Species: Candida albicans

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Mar-2000

C:Accession: T30531

R:Gaur, N.K.; Klotz, S.A.

Infect. Immun. 65, 5289-5294, 1997

A:Title: Expression, cloning, and characterization of a Candida albicans gene, ALA1, the

A:Reference number: 220847; PMID:98053977; PMID:9393828

A:Accession: T30531

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1419 <GAU>

A:Cross-references: EMBL:AF025429; NID:G2522210; PID:G2522219; PIDN:AA888883.1

C:Genetics:

A:Gene: ALA1

C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

```
Query Match 49.2%; Score 3194; DB 2; Length 1419;
Best Local Similarity 49.5%; Pred. No. le-130;
Matches 708; Conservative 182; Mismatches 355; Indels 186; Gaps 29;

QY 1 MLOQFTLLFLYLSIASAKTITGVDFSNLSLWNAANYAFKPGCYPTWNAVLGWSLDGTS 60
|||||
Db 1 MLOQFTLLFLYLSIASAKTITGVDFSNLSLWNAANYAFKPGCYPTWNAVLGWSLDGTS 60
QY 61 ANPGDFTLLNMPGVFKYTTTSOTSDLTADGVKATCOFYSGEEFTTSTLCTVNDALKS 120
|||||
Db 61 ANPGDFTLLNMPGVFKYTTTSOTSDLTADGVKATCOFYSGEEFTTSTLCTVNDALKS 120
QY 121 SIKAFGTVTLPIAFNVGGTSGTDLSDSKCFTAGTNTVTFNDGSKLSIAVNFEEKSTVDP 180
|||||
Db 121 SIKAFGTVTLPIAFNVGGTSGTDLSDSKCFTAGTNTVTFNDGSKLSIAVNFEEKSTVDP 180
QY 181 SAYLYASRVMPSLNKVTTLEVAPOCENGYSITGTMGSSNGDVAIDCSNTHIGITKGLND 240
|||||
Db 181 SAYLYASRVMPSLNKVTTLEVAPOCENGYSITGTMGSSNGDVAIDCSNTHIGITKGLND 240
QY 181 SGYLTTSRFPLSKATLTVAPQCENGYTSMTGFSTSGYDVAIDCSNTHIGISKVND 240
|||||
Db 181 SGYLTTSRFPLSKATLTVAPQCENGYTSMTGFSTSGYDVAIDCSNTHIGISKVND 240
QY 241 WNPVSVSESPSYTKTCTSGNIGIQKYNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
|||||
Db 241 WNPVSVSESPSYTKTCTSGNIGIQKYNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
QY 301 SRLQSKPFTLRWTKYKNSDAGSNGIVVATRTVDTSTTAVTTLPPNPSVDKTKTIELQ 360
|||||
Db 301 SRLQSKPFTLRWTKYKNSDAGSNGIVVATRTVDTSTTAVTTLPPNPSVDKTKTIELQ 360
QY 361 PIPTTTTTSYGVVTSYLTKTAPIGETATVIVDVPYHTTITVISEWGTITTTTNTNP 420
|||||
Db 361 PIPTTTTTSYGVVTSYLTKTAPIGETATVIVDVPYHTTITVISEWGTITTTTNTNP 420
QY 421 TDSIDVWVQVPLNPNTVSTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEY 480
|||||
Db 421 TDSIDVWVQVPLNPNTVSTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEY 480
QY 481 SQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVPAPGGTDSVIRE 540
|||||
Db 481 SQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVPAPGGTDSVIRE 540
QY 541 PNPVTTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 600
|||||
Db 541 PNPVTTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 600
QY 601 PPGTDSVIREPPNPVTTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEY 657
|||||
Db 601 PPGTDSVIREPPNPVTTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEY 657
QY 658 -----EYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 691
|||||
Db 658 -----EYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 691
QY 661 SSSDNISSSAQSSSSVEQSSIVGLSSSDIPLSDMPSSSSSTGLTSSSSTVSSYDSD 720
|||||
Db 661 SSSDNISSSAQSSSSVEQSSIVGLSSSDIPLSDMPSSSSSTGLTSSSSTVSSYDSD 720
QY 692 -----TTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 725
|||||
Db 692 -----TTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 725
QY 721 SSSSIESSTLSSDRCSSISDITTSFWDSSSDLESTITWSSSIDAQSSHLVQSVNSI 780
|||||
Db 721 SSSSIESSTLSSDRCSSISDITTSFWDSSSDLESTITWSSSIDAQSSHLVQSVNSI 780
QY 726 VTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 768
|||||
Db 726 VTTTEYWSQSFATTTTVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 768
QY 781 STSQELSSSSSESSSTF-----ATDALVSSDASSILLSSDTSYYPSTISSDDFPHIA 835
|||||
Db 781 STSQELSSSSSESSSTF-----ATDALVSSDASSILLSSDTSYYPSTISSDDFPHIA 835
QY 769 -----SQSFATTTT-----TVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 806
|||||
Db 769 -----SQSFATTTT-----TVPAPGGTDSVIREPPNPVTTTTEYWSQSFATTTTVA 806
QY 836 GESDLSISIPITSTVEITSDVSLTSDPASFFSSSSSLNDSNDSPPSSDQSDILTSSSFS 895
|||||
Db 836 GESDLSISIPITSTVEITSDVSLTSDPASFFSSSSSLNDSNDSPPSSDQSDILTSSSFS 895
QY 807 TSIIIPSF-----RPHYVNSYTSDLSTPSSSMNPTSISSDGLMLSLTTLVTS 856
|||||
Db 807 TSIIIPSF-----RPHYVNSYTSDLSTPSSSMNPTSISSDGLMLSLTTLVTS 856
QY 896 TLVVPFSLSSSSSSLTTPHYVNSYTHASESESSSVASPSMASAN--DDTYLTSES 953
|||||
Db 896 TLVVPFSLSSSSSSLTTPHYVNSYTHASESESSSVASPSMASAN--DDTYLTSES 953
QY 857 ETTTELICSDGKE---CSRLSSSSSGIVTN-PDNESSIVTSTVPTASTMSDLSSTDG-I 911
|||||
Db 857 ETTTELICSDGKE---CSRLSSSSSGIVTN-PDNESSIVTSTVPTASTMSDLSSTDG-I 911
QY 954 TDTTSSIGTSSVTCRRDNGDCIVTGMPPSSSIDSEQTSDVTTTSSFFVASSTPTSAEQ 1013
|||||
Db 954 TDTTSSIGTSSVTCRRDNGDCIVTGMPPSSSIDSEQTSDVTTTSSFFVASSTPTSAEQ 1013
QY 912 SATSSDNV--SKSGVSVTTTSTVTTTQTPNPLSSSVTSLTQLSSPSVSESESKVFTTS 969
|||||
Db 912 SATSSDNV--SKSGVSVTTTSTVTTTQTPNPLSSSVTSLTQLSSPSVSESESKVFTTS 969
```

Db	259	SQPNCTIIPDPSNVTVSTIITTTB	-----WTGFTTSTSTEMTVTGTNGVDPDDEVIVIRT	314
Qy	332	RTVTDSTTAVTTLLPNPNSVDKTKT	-----IEILQP-IPtTITTT	369
Db	315	PT-TASTIITITTEPWNSTFTSTSTELT	TVGTNGVRDDETIIVIRPTTATTAITTTPEW	373
Qy	370	--SYVGVTTSYLKTA	---PIGETATVIDVVPYHHTT--TVISEWTGTTT	415
Db	374	NSTFTSTSTELT	TVGTNGLPDDE-TIIVIRPTTATTAITTTQPMWDTFTSTSTELT	432
Qy	416	TRTN--PTDSIDTVVVOVPLNPNT	-VSTTEWWSQSFATT---TTVTAPPG-GTD--TVI	465
Db	433	TGNGLPDDE-TIIVIRPTTATTAITTTQPMWDTFTSTSTELT	TVGTNGLPDDETIIVIRPTTATTAITTTQ	491
Qy	466	IREPPNHT--VTTTEYWSQSFATT	---TTVTAPPG-GTDS--VIREPPNPT--VTTE	514
Db	492	IRPTTATTAMTTQPMWDTFTSTSTELT	TVGTNGLPDDETIIVIRPTTATTAITTTPEWNTFTSTSTEMTT	551
Qy	515	YWSQSFATT	---TTVTAPPG-GTDS--VIREPPNPT--VTTTEYWSQSYATT	561
Db	552	PWNDFTSTSTEMTVTGTNGLPDDETIIVIRPTTATTAITTTPEWNTFTSTSTEMTT	611	
Qy	562	VTAPPG-GTDS--VIREPPNHT--VTTTEYWSQSYATT	---TTVTAPPG-GTD--TVI	609
Db	612	VTGNGLPDDETIIVIRPTTATTAITTTQPMWDTFTSTSTEMTVTGTNGLPDDETIIV	671	
Qy	610	IREPPNHT--VTTTEYWSQSFATT	---TTVTGP---PSGTDTVIIREPPNPT--VTTE	658
Db	672	IRPTTATTAMTTQPMWDTFTSTSTELT	TVGTNGLPDDETIIVIRPTTATTAITTTQ	731
Qy	659	YWSQSYATT	---TTITAP---PGEDTVIIREPPNHTV--TTEYWSQSYATT	705
Db	732	PWNDFTSTSTEMTVTGTNGVPDDEVIRPTSEGLTITTEPEWNTFTSTSTEMTT	791	
Qy	706	VTAPPGE-TD--TVIIREPPNH--TVTTEYWSQSYATT	---TTVTAPPG-GTD--TVI	753
Db	792	VTGNGOPTDDEVIRPTSEGLTVTTTEPEWNTFTSTSTEMTVTGTNGVPDDEVIV	851	
Qy	754	IREPPNHTV--TTEYWSQSFATT	---TTVTAPPG-GTDTVIIESSMS---SSKISTS	802
Db	852	IRPTSEGLTITTEPEWNTFTSTSTEMTVTGTNGOPTDDEVIRPTTATTAISSLS	911	
Qy	803	SNDTIIPS---FSRPHYVNSTSDLTSESSMNPTSISSDGM	---LLSSTLVTES	856
Db	912	SGQITSFITSARPIITPEYPSNGTSVISSSVISSSDTSLLVISSVTSLLVSSPVISS	971	
Qy	857	ETTELLCSGKGCBSRLSSSGVITNPDNSNESSIVTSTVPTASTWSDLSLSTGSIATSS	916	
Db	972	FISSPVSSITTSASILSESS-----KSSVITPSTSGSSESTGCSASS	1018	
Qy	917	DNYSKGVSVTFTSVTTIOTTNPPLSSSVTSLTQLSSIPSVSESEKVTFTSNGDNOSG	976	
Db	1019	---SSSISSESPKSTVSSSLP-PVTSATTSQBITSLLPVTTKT	1061	
Qy	977	THDSQSTSTEIEIVTTSSTKVLPPVSSNDLTASEPT	---NTRQOPTTLSTSTNSIETDI	1033
Db	1062	-----SQOTLLVTVTSCESHVCTESISSAIVSTATVTVSGATTEYTWCPITSTEITKQT	1116	
Qy	1034	T-TSQPGDNGDNTSSNPVPTVATSTLASASED	--NKSGSIESASTSLKPSMGENSEGL	1090
Db	1117	TETTKQKGTGEQTETTKQTVVVT	---ISSCESDVCSKTASPAIVSTATNGVTVTEY	1173
Qy	1091	TTSTEIEATTSPTPEAPSPAVSGDVTTEPTDTRQOPTTLSTTSKTNSELVAT	---T	1145
Db	1174	TTWCPISTTESKQOTLLVTVTSCSGVCSFT	-----SPAIVSTATATVNDVVVYVSWR	1229
Qy	1146	QATNENGGKSPDTLTSSLTGTGSAST	-----SANSSELTVSGSVT-----GGAVASASN	1194
Db	1230	QTTNEQSVSSKMNASATSETTNTGAAETTTSTGAAETKTVVTSISRENHAETOTASATD	1289	
Qy	1195	DQSHSTSV-----TNSNSIVSNPTQTLTSSQVTS	-----SPSTNTFIASTYDGSII	1243
Db	1290	VIGHSSVWSVSETGNTKSLTSSGLSTMSQPRSTPASMMVGSGTASLESTYVAGSANS	1349	

	Query Match	14.5%	Score 939;	DB 2;	Length 1367;		
	Best Local Similarity	28.1%	Pred. No. 2.9e-33;				
	Matches	410;	Conservative	214;	Mismatches	469;	
				Indels	364;	Gaps	84;
QY	28	NSLTWSNAANYAFKGGPYTNNAVLGWSLGGTSANPGDFTFLNMPGVKVTTSQT-----S	83				
DB	50	DSSTYSNAAYMAY---GYASKT-----KLGSGVGQTDISIDYNIPCV---SSSGTFPCPQ	98				
QY	84	VLDTAD-GVK-YATCO-----FYSGEEFTTFSTLCTVNDALKSSIKAKGTVTLPTA--	133				
DB	99	EDLYGNMGCKGIGACSNPIIAYWSTDLFQFTPT-----NVTLEMTGY	143				
QY	134	FNVGCTGSSDLEDSKCFTAGTNTVTNDCGDKDISIDVEFEKSTVDPSPAYLYASRVMPSL	193				
DB	144	FLPQTGSGYT-----FKFATVDDSAIL-----	165				
QY	194	NKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDC-----SNIHIGITKGLNDWN-	242				
DB	166	-----SVGGSIAFECCAQEQPPISTNFTI---NGIRPWNG	198				
QY	243	YPVSESFSYTKTCTSNGIQIKYQNVPY-GYRPF-----IDAYISATDVN-	286				
DB	199	SPPDNIITGVYMYAGFYIPMKIYVSNVANGTLPISVTLPDGTVSDDFEGYVYTFDNNL	258				
QY	287	---QYTLAYVNDYTCAGSRQSLKFTLRWTGYNKSDA-----GSGNGI-----VIVATT	331				

QY 1244 QHSTWLYGLTLLSLFI 1260
 Db 1350 LAGSGLSVFIASLLAI 1366

RESULT 4
 S53465
 flocculation protein FLO1 precursor - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein YAR050w
 C:Species: Saccharomyces cerevisiae
 C:Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999.
 C:Accession: S53465; S43543; S31230; S57349
 R:Bussey, H.; Keng, T.; Storms, R.K.; Vo, D.; Zhong, W.; Fortin, N.; Barton, A.B.; Kabac
 submitted to the EMBL data library, February 1994
 A:Description: Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of the 5
 A:Reference number: S53458
 A:Accession: S53465
 A:Molecule type: DNA
 A:Residues: 1-1537 <BUS>
 A:Cross-references: EMBL:L28920; NID:g1616966; PIDN:AAC09499.1; PID:g694125; MIPS:YAR050
 R:Watarai, J.; Takata, Y.; Ogawa, M.; Sahara, H.; Koshino, S.; Onnala, M.L.; Airaksinen,
 Yeast 10, 211-225, 1994
 A:Title: Molecular cloning and analysis of the yeast flocculation gene FLO1.
 A:Reference number: S43543; MUID:94262325; PMID:8203162
 A:Accession: S43543
 A:Molecule type: DNA
 A:Residues: 1-428, 'M', 430-473, 'M', 475-518, 'M', 520-549, 'T', 551-608, 'L', 610-636, 'M', 638-69
 R:Onnala, M.
 submitted to the EMBL data library, September 1994
 A:Reference number: S57851
 A:Accession: S57851
 A:Molecule type: DNA
 A:Residues: 1-428, 'M', 430-463, 'D', 465-473, 'M', 475-518, 'M', 520-549, 'T', 551-608, 'L', 610-63
 A:Cross-references: EMBL:X78160; NID:g535933; PIDN:CRA55024.1; PID:g535934
 R:Teunissen, A.W.R.H.; Holub, E.; van der Hucht, J.; van den Berg, J.A.; Steensma, H.Y.
 Yeast 9, 423-427, 1993
 A:Title: Sequence of the open reading frame of the FLO1 gene from Saccharomyces cerevisi
 A:Reference number: S31230; MUID:93289821; PMID:8511970
 A:Accession: S31230
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-296, 927-1516, 'TAWPVVV' <TEU>
 R:Bidard, F.; Bony, M.; Blondin, B.; Dequin, S.; Barre, P.
 Yeast 11, 809-822, 1995
 A:Title: The Saccharomyces cerevisiae FLO1 flocculation gene encodes for a cell surface
 A:Reference number: S57349; MUID:96090130; PMID:7483845
 A:Accession: S57349
 A:Molecule type: DNA
 A:Residues: 1243-1274; 1308-1339; 1359-1390 <BID>
 C:Genetics:
 A:Gene: SGD:FLO1
 A:Cross-references: SGD:S0000084; MIPS:YAR050w
 A:Map position: 1R
 C:Keywords: duplication; glycoprotein; transmembrane protein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-1537/Product: flocculation protein FLO1 #status predicted <MAT>
 F:953-997/Domain: repeat A2 <RA2>
 F:998-1042/Domain: repeat A3 <RA3>
 F:1043-1081/Domain: repeat A4 <RA4>
 F:1226-1276/Domain: repeat B1 <RB1>
 F:1277-1284/Domain: repeat B2 (partial) #status atypical <RB2>
 F:1291-1341/Domain: repeat B3 <RB3>
 F:1342-1392/Domain: repeat B4 <RB4>
 F:1408-1416/Domain: repeat C1 <RC1>
 F:1417-1425/Domain: repeat C2 <RC2>
 F:1426-1434/Domain: repeat C3 <RC3>
 F:135,187,262,1114/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.5%; Score 939; DB 2; Length 1537;

Best Local Similarity 25.8%; Pred. No. 3.4e-33;

Matches 414; Conservative 246; Mismatches 520; Indels 426; Gaps 84;

QY 2 LQOFTLLFLYLSIASAKT-----ITGVFDSF-----NSLTWSNAANYAF----- 40
 Db 10 LAVFTLLAL--TSVASCATEACLPAGQKSGMNFQYSLKDSSTYSNAYMAYGYASKT 68
 QY 41 -----KGPGYPTW-----NAVLGWSLD--G 58
 Db 69 KLGSGVGQTDISIDYNIPCVSSSGTTPCQEDSYGNWCKGKMGACSNQGIAYWSTDLFG 128
 QY 59 TSANPGDTFTLNM-----PCVFKYTTSO-----TSVD 85
 Db 129 FYTPTTNV-TLEMTGYFLPPQTGTYTFKFATVDSDSAILSVGGAFAFNCACCAQQOPIITSTN 187
 QY 86 LTADGVK-----YAPCQYSGEEFTTSTLCTVNDALKSSSKAFGT-----VTLP 131
 Db 188 FTIDGKPKWGSGLPPNIEGTVMYAGYYPM-----KVYSNANVSGTLPISVTLTP 238
 QY 132 IAFNVGGTSGSTDLEDKSCFTAGTNTVFNDDKDISIDVEFEKSTV-DPSAYLYASRVM 190
 Db 239 -----DGTTVSDDFEG-----YVVSFDD-----DLSQSNCTVPDPSNTA-VSTTT 277
 QY 191 PSLNKVTTTLFVAPQCENGYSYTGTMGSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESF 250
 Db 278 TTTEPWTGTFSTSTEMTTVTGNGVPTDETIVIRTPPTASTIIITTEPNWSTFTSTST 337
 QY 251 SYTKTCTSGIQIKYQNVAPAGYRPFIDAYISATDVNOYTLAYTNDVTCAGSRLOSKRPFTL 310
 Db 338 ELTIVTGTNGVRDEII-----VIRTPPTATTAITTEPNWSTFTSTSTSTEL----- 384
 QY 311 RWTGYKNSDAGSNGI---VIVATRTVDTOSTTAVTTL-PENPSVDKTKT----- 355
 Db 385 -----TTVTGTLNGLPTDETIIVIRPTTATTATMTTQPNWDTFTSTSTELTGTGNGL 438
 QY 356 -----IEILOPPTTIIT-----SYGVVTSYLTAKTA---PIGETATVIVDVPYH 398
 Db 439 PDETIVIRTPPTTATTATMTTQPNWDTFTSTSTELTGTGTLNGLPTDET-IIIVIRPTT 497
 QY 399 TTT--TVTSWTCITII-----TTTTRTN--PTDSIDTVVQVPLPNPTVSTT-----EY 443
 Db 498 ATTATMTTQPNWDTFTSTSTELTGTGTLNGLPTD--ETIIV---IRPTTATTAMTTPQ 552
 QY 444 WSQSFAAT---TTVTAPPG-GTD--TVIIREPPNHT--VTTEYSQSFAAT---TTV 490
 Db 553 WNDFTSTSTEMTTVTGTLNGLPTDETIIVIRTPPTATTATTAITTEPNWSTFTSTSTEMTTV 612
 QY 491 TAPPG-GTDS--VIREPPNPT--VTTEYSQSFAAT---TTVTAPPG-GTDS--VII 538
 Db 613 TGTNGLPTDETIIVIRTPPTATTATMTTQPNWDTFTSTSTEMTTVTGTLNGLPTDETIIVI 672
 QY 539 REPPNPT--VTTEYSQSFAAT---TTVTAPPG-GTDS--VIREPPNHT--VTTEY 587
 Db 673 RTPTTATTATMTTQPNWDTFTSTSTELTGTGTLNGLPTDETIIVIRTPPTATTAMTTPQ 732
 QY 588 WSQSFAAT---TTVTAPPG-GTD--TVIIREPPNHTV--TTTEYSQSFAAT---TTV 634
 Db 733 WNDFTSTSTEMTTVTGTLNGLPTDETIIVIRTPPTSEGLISTTTEPWTGTFSTSTEMTTV 792
 QY 635 T---GPSSGTDVIIIREPPNP--TVTTEYSQSFAAT---TTIAP---PGETDVLII 682
 Db 793 TGTNGLPTDETIIVIRTPPTSEGLISTTTEPWTGTFSTSTEMTTVTGTLNGLPTDETIIVI 852
 QY 683 REPPNHTV--TTTEYSQSFAAT---TTVTAPPG-GTD--TVIIREPPNHTV--VTTEY 731
 Db 853 RTPTSEGLISTTTEPWTGTFSTSTEMTTGTGTLNGLPTDETIIVIRTPPTSEGLISTTTEP 912
 QY 732 WSQSFAATTTVTAPPGTD-----TVIIREPPNPTV--TTTEYSQSFAATTTVTAPP 782
 Db 913 WTGTFSTSTEMTHVTGTLNGLPTDETIIVIRTPPTSEGLISTTTEPWTGTFSTSTSTETI 972
 QY 783 GGT-----DTVIIYESMSSSKI-----STSSNDITSIIIPFSRPP----- 816
 Db 973 TGTNGLPTDETIIVIRTPPTSEGLISTTTEPWTGTFSTSTEMTTVTGTLNGLPTDETIIVI 1032

Db	178	G-----CNNVD---NQCHSQTDFP-----GFYWNIDCDNN-----C 205
Qy	281	SATDVNOYTILAYTNDYTCAGSRQLSQKFTLRWTKYKNSDAGSGIVAVTRTVTDSSTA 340
Db	206	GGRKSSITTSSTSESSITTSTSESSITT-----SSTSESS-----TTTSSTSESSSTS 254
Qy	341	VTLPLFPNPSVDKTKIIELOPIPTTTTTSYGVVTTSLYKTAPIGETATVINDVPXYHT 400
Db	255	STTAPATPT-----TTSCTEKEPPPTTIS-----CTKEKP-----TPPHDHT 292
Qy	401	TTVTSEWTGTTTTTTTN--PIDSIDTV-VVOVPLPNPTSVSTXYSQ--SFATTTV 454
Db	293	TPCTKKKTKTKTKTKTKTTVPVPSSTSESSAPVTPSSSTSESSAPVTSSTSESS 352
Qy	455	TAP----PGGDTVIIRPPNHVTTTYEWSQFATTTVTAPPGGTDSVIIREPPNPVT 511
Db	353	SAPVPTSSSTSESSAPVTSSTSESSAPVTSSTSESSAP-----VPTSSS 401
Qy	512	TTYEWSQFAATTTVTAPPGGTDSVIIIREPNPTVTTTYEWSQYATTTT-VTAPPGGT 570
Db	402	TTESSAPVTSSTSESSAPVTSSTSESSAPVTSSTSESSAPVTSSTSESSAPVTP 461
Qy	571	SVIIREPNHVT--TTEYS-----OSYATTTVTAPPGGTDVIIREP-PNHTVTTE 622
Db	462	SSSTSESSAPVTSSTSESSAPVTPSSSTSESSAPVTSSTSESSAPVTPSSSTSE 521
Qy	623	YWS-----QSFAATTTVTGPPSGDTHIIREP-PNPTVTTTYEWSQYATTTITAPGE 676
Db	522	SSSAPATPSSSTSESSAPVTSSTSESSAPVTPSSSTE-----SSPTVSSTTE 575
Qy	677	TDTHLIREPNHVTTTYEWSQYAT---TTTVTAPPGETDVLIREPNHVT--TTE 730
Db	576	SSSAPV--PTPSSSTSESSAPVTPSSSTSESSAPVTPSSSTSESSAPVTSSTTE 632
Qy	731	YWS-----QYATTTVTAP---PGGDTVIIREP-PNPVTTTYEWSQFATTTVTAP 781
Db	633	SSSAPVTPSSSTSESSAPVTPSSSTSESSAPVTPSSSTTE---SSAPVTSSTTES 690
Qy	782	PGGDTVIIYESMSS---SKISTSNIDITSIIIPFSR-----PHYVNSTSD--- 825
Db	691	SSAPVTSSTSESSAPVTPSSSTSESSAPVTPSSSTSESSAPVTPSSSTSESSA 750
Qy	826	---LSTPESSSMNPTSISSDGMLLSSTLVITESETTELICSDGKCRLSSSSGIVTN 882
Db	751	PVTSSTTESSAPVTP-SSSTTESSAPVTPPSSSTTESSAPVTPPSSSTTESSAPV 809
Qy	883	PDSNESSIVTPTASTMGDSL--STDGIGATSDDNVKSQSVVTTETSVTTIQTPNP 941
Db	810	PTSSSNITSSAPSPFPSSSTESSVPVTPPSSSTSESSAPVSSSTTESSVAPVTP 869
Qy	942	LSSSVTSLTOLSSIPTSVESES-----KYFTSGDNQSGTHDSQSSTLEIV----- 990
Db	870	SSSNITSSAPSSIPPSTTESFSTGTVTVPSSSKYPGSTQETSSTETTIVPTKTT 929
Qy	991	---TTSSTKVLPPVWSNTDLTSEPTN-----TREOPTLTST-TSNSITEDITHSQ 1038
Db	930	SVITPSTTTTTVCSTGCTNSAGETTSGCSPKIVITVTPTTTTTSVTSTSTTTITTVCS 989
Qy	1039	TGDN--GDNTSSTNPVPTVAITSLTASAEEDNKSGSHESA-----SST 1078
Db	990	TGNTSAGETTSGCSP-KTIITTVPCSTSPSETASESTTSPPTPVTTWSTVVTVTEYST 1048
Qy	1079	SLRPSMENGSLTIS--TEIEANTTSPEAPSVA-----SSGTD----- 1116
Db	1049	STRPG-GE---ITTFVTKNIPTTYLIAPTTSVTVTNFTPTITTVCSSTGCTNSAGE 1104
Qy	1117	-----VTEBPTOT-----REOPTLTSTSKTSELNAVTAQATNGKGKSPSDLT 1161
Db	1105	TTSCSPKVTTTVPCTGCTGEYTEATTLTVAVTIT-VVTESSTGCTNSAGTKTTGYT 1163
Qy	1162	SSLTGTGTSANSSELVTSG-----SVTGGAVASANDQOSHSTSVTNSNISVN---- 1211
Db	1164	TKSVPTTIVTTLAPSAVPTPATINAVPTIITTECSAATNAAGETTSVCSAKTIVSSASG 1223

Db 2873 TATTATGCTATPSSPTGCTGPPKVLTPATTTATSSKATSSSPRTAT-TLP----- 2926
 QY 571 SVIIREPPNHTVT-TEYWSQSYATTTV-----TAPGGCTDVTIIEPPNHTVTTEYWS 625
 Db 2927 -VLTSTAKSTATSVPIPSSTLGTGTLPEQTTTPVAMSTIHPSTPETHTSTVLTT 2985
 QY 626 QSFATTTVTPGSPGTDVLIIEPPNPTVTTEYWSQSYATTTTITAPPGCTDVTIIEP 685
 Db 2986 KATTRATSSSTPSSST-----PGTWILTE-----LTTAATTTAGTCPTAT----- 3026
 QY 686 PNHTVTTEYWSQSYATTT-TVTAPGCTDVTIIEPPNHTVTTEYWSQSYATTTTVA 744
 Db 3027 PPSPTGCT--WILTELTATTTASTGSTAT--LSSTPGCTWILTE-----PSTTATVA 3077
 QY 745 PPGCTDVTIIEPP--NPTVTTEYWSQSYATTTTVA---PPGCTDVTIIESSSSK 798
 Db 3078 PGSTATASQTATAGTGHVSTT-----ATTPTVTSSKATPSSSPGCTATALPAASTA 3130
 QY 799 ISTSSNDITSIIPFGRPHVYNSITDLS--TFESSMNTPTSISSDGLMLLSSTTLVTES 856
 Db 3131 TTPATATFTA-IPSSS-----LGTWTRLSQTTTPATMTATPSSSTPETVHTSTVLTTTA 3185
 QY 857 ETTELICSDGKCSRSSSSGIVTNPDSNESSIVTPTASTMSDLSLSDGICATSS 916
 Db 3186 TTT-----GATGSVATPSSPTGTAHTTKVPT-----TTTGTATPS 3222
 QY 917 DNVKSGSVT-----TETSVTTIQTTPNPLSSSVTSLTOLSIPIVSSESKVTFTSNGD 972
 Db 3223 ---SSPGTALTPPWISTTTPPTTTPPTGSGVTP-----SSIPGT----- 3261
 QY 973 NQSGTHDSQSTFEIIVTSSTKVLPPVYSSNTDLTSEPTNTREQPTTLSTNSITED 1032
 Db 3262 ---TUTAR-----VLTTT---TVATGSMATPSSSTQSTGPPPSLTATTATITAT 3306
 QY 1033 ITTSQPTGNDGNTFSNPVPTVATSTLASASEDNKSHESASLSKPSMGENGSLT 1092
 Db 3307 GSTNPS-----STPGCTPIPPVLTSMTATPAATSSK-----ATSSSPRTATLPLVLT 3355
 QY 1093 STEIATTSPTAPSPAVSSGTDV---TTEPTDTRTQPTTLSTKTNSELVATTOQTN 1149
 Db 3356 STAKSTATSTFTPIPSSTLWTTVPAQTTTPMTSTHTSTPETHTSTVLTTTATM 3415
 QY 1150 ENGKSPSDDLSTLGTGSASTSANSSELVTSQSVTGGAVASANDQSHSTS----- 1201
 Db 3416 T---RATNSTATPSLTGT---FRILTELTATTTAATGSTATLSLSTPGTWTILPEPT 3469
 QY 1202 -----VTNSIVSVNTPTTLTSSQVTSPPSTNTFIASYDGSST 1242
 Db 3470 IATVMVPTGSTATSTLGTGTAHTPKVVTAMTMTATATVPSSSTV 3516
 RESULT 11
 A43932
 mucin 2 precursor, intestinal - human (fragments)
 N:Alternate names: mucin 5Muc-41
 C:Species: Homo sapiens (man)
 C>Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
 C:Accession: A49963; A45106; A43932; B33532; A61257; PQ0329
 R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
 J. Biol. Chem. 269, 2440-2446, 1994
 A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
 A:Reference number: A49963; MUID:94132002; PMID:8300571
 A:Accession: A49963
 A:Molecule type: mRNA
 A:Residues: 1-639 <GU1>
 A:Cross-references: GB:L21998
 R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
 J. Biol. Chem. 267, 21375-21383, 1992
 A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
 A:Reference number: A45106; MUID:93016075; PMID:1400449
 A:Accession: A45106
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA

A:Residues: 626-1895 <GU2>
 A:Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
 A:Note: sequence extracted from NCBI backbone (NCBIP:116706)
 A:Accession: B45106
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 2037-3020 <GU3>
 A:Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
 A:Experimental source: colon
 A:Note: sequence extracted from NCBI backbone (NCBIP:116698)
 R:Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
 J. Clin. Invest. 88, 1005-1013, 1991
 A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
 A:Reference number: A43932; MUID:91358717; PMID:1885763
 A:Accession: A43932
 A:Molecule type: DNA
 A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>
 A:Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
 A:Note: sequence inconsistent with the nucleotide translation
 A:Note: sequence extracted from NCBI backbone (NCBIP:55750)
 R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.
 J. Biol. Chem. 264, 6480-6487, 1989
 A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
 A:Reference number: A33532; MUID:89197956; PMID:2703501
 A:Accession: B33532
 A:Molecule type: mRNA
 A:Residues: 1916-2193 <GU4>
 A:Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
 A:Experimental source: intestine
 R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
 J. Clin. Invest. 87, 77-82, 1991
 A:Title: Human bronchus and intestine express the same mucin gene.
 A:Reference number: A61257; MUID:91086481; PMID:1985113
 A:Accession: A61257
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>
 A:Experimental source: bronchus
 R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstn
 Biochem. Biophys. Res. Commun. 183, 821-828, 1992
 A:Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the
 A:Reference number: PQ0328; MUID:92198477; PMID:1550588
 A:Accession: PQ0328
 A:Molecule type: mRNA
 A:Residues: 2328-2468 <XUG>
 A:Cross-references: GB:M86523
 A:Experimental source: small intestine
 A:Accession: PQ0329
 A:Molecule type: protein
 A:Residues: 2328-2342, 'K', 2344-2354 <XUG1>
 C:Genetics:
 A:Gene: GDB:MUC2
 A:Cross-references: GDB:L20203; OMIM:158370
 A:Map position: lip15.5-lip15.5
 C:Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; v
 C:Keywords: glycoprotein; intestine; tandem repeat
 F:2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>
 Query Match 9.2%; Score 599; DB 2; Length 3020;
 Best Local Similarity 25.8%; Pred. No. 3.9e-18;
 Matches 285; Conservative 120; Mismatches 438; Indels 262; Gaps 44;
 QY 249 SFSYTKTCTSNGLQIKYQNV-----PAGYRPFIDAYISATDVNOYTLAVTNDYTCAGSR 302
 Db 1236 AFCYWEICGNGTVEKHFNCISITTRPSLTITFTTLTPPS-FTTTTTTTPSSTV 1294
 QY 303 LQSKP-FTLRWTGYKNKSDAGSNGI-----VIVA-----TTRTVDSTTAVT----- 342
 Db 1295 LSTTPKLCCLWSWDINEDHPSSGSDGDRPDPGVCAPEDICRSVKDPHLSLEQHGOK 1354
 QY 343 -----TLFPNPSYDKTKTIE-----ILQIPPTTTTITSYGVVIT 376
 Db 1355 VOCDSVSGFICKNEQDGFGLCYDKIRVNCNCPMDKCIITPSPPTTTPSPPTTTT 1414


```
Db 813 EIVITSCSKSCTNEDSVLTQVSTVETITSC-----SGICITLMSPVTTINAKANT 866
Qy 1132 STTSKSELVATQATNENGKGS-----PSTDLTSLTGTASPSANSELVTS----- 1181
Db 867 LTTTETS--VETTTTCGGVCSLTVVPVTTITSEATTATATISCEDNEEDITSTETELL 924
Qy 1182 -----GSGVTGAVASAN-----DQSHSTSVTNSNSIVSNTPQ-----TLQQQVTS 1223
Db 925 TLETTITSCGGICITLMSPVTTINAKANTLTITTSVEITITTCSSGGVCSLTVVPVT 984
Qy 1224 -SSPSTNTFIASDTGSGSIHQHSTWLYGLITLLS 1257
Db 985 ITSEATTATATISCEDNEEDVASTKTELLMETTIT 1019

RESULT 14
T32271
Hypothetical protein ZC178.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32271
R:Murray, J.; Wohldmann, P.; Beck, C.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid ZC178.
A:Reference number: Z21143
C:Accession: T32271
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1459 <MUR>
A:Cross-references: EMBL:AF024496; PIDN:AAB70340.1; GSPDB:GN00023; CESP:ZC178.2
A:Experimental source: strain Bristol N2; clone ZC178
C:Genetics:
A:Gene: CESP-ZC178.2
A:Map position: 5
A:Introns: 1099/2

Query Match 8.5%; Score 553.5; DB 2; Length 1459;
Best Local Similarity 24.5%; Pred. No. 1.5e-16;
Matches 333; Conservative 172; Mismatches 593; Indels 261; Gaps 55;

Qy 57 DGTANPGDTFT-----LMPGVKTYTTGTSVLDLTADGVKAT----- 95
Db 158 DSTAGPSSTATNSASSETPC-----NSETQ-----TSDGTSTMTVPNDSTTAGPSSTATNS 209
Qy 96 -----COFYSGEETFTSTLCTV-NDALKSSIKAFGTVLPIAFNVCGTSGSTDLEDS 148
Db 210 ASSETPC-----NSETQSDGTSTMTVPNDST-----TTAGPSSTATNSASSETPCNSE 257
Qy 149 KCFTAGTNTVTFNDGDKDITSDVEPEKSTVDPSAYLYASRVNPKVTKVTLFVAPQCENG 208
Db 258 TOTSDGTSTMT-----VPNDSTAGPS-----STATNSASSETPCNSETOTSDG 301
Qy 209 YTSGTMGTFSSNGDAIDCSNIHIGITKGLNDWNPVSESPSYTKTCTNSG---IQIKY 265
Db 302 TSTMTVPNDSTTAGPSSTATN-----SASSETPCNSETOTSDGTSTMTVPN 347
Qy 266 QNVGAGYRPFIDAYISA-----TDVNQYTLATNDYTCAGSRLOKRPFLRWTC 314
Db 348 DSTTAG--PSSTATNSASSETPCNSETOTSDGTSTMTVPNDSTTAGPSSTATNSASSETP 405
Qy 315 YKNSDAGSNGVIVATRTV--TDSTTA-----VTF-----LPFNPSV---DKTKTIEI-- 358
Db 406 CNSETQTSOG-----TSTMTVPNDSTTAGPSSTATNSASSETPCNSETOTSDGTSTMTVSS 461
Qy 359 --LQPIPTTITTSYGVVT-----TSLYTKTAP-----GETATVIVDPVYHT--- 399
Db 462 DSTTGPCSSTATNSASSETPCNSETOTSDGTSTMTVPNDSTTAGPSSTATNSASSETPCN 521
Qy 400 TTTVTSWTGTTTTTTTTRNPTDSIDTVVQVPLNPNVSTTEYWSQSFATTTVTAP--- 457
Db 522 SETQTSOGTSTMTVSSDSTTGPSS--STATNSASSETPCNSE-----QTSOGTSTMTVPND 576
```

```
Qy 458 -----PGCDDT--VIREPPNHTVTTTEYWSQSFATTTTIVTAPPG--TDSVIREPPNP 508
Db 577 STTAGPSSTATNSASSETPCNSETOTSDGTSTMTVPNDSTTAGPSSTATNSASSETPCNS 636
Qy 509 TVTTTEYWSQSFATTTTIVTAPPG--GTDSVIREPPNHTVTTTEYWSQSFATTTVTAP 566
Db 637 ETQTSOGTSTMTVSSDSTTAGPSSTATNSASSETPCNSETOTSDGTSTMTVSSDSTTAG 696
Qy 567 G--GTDSVIREPPNHTVTTTEYWSQSFATTTTIVTAPPGTDT--VIREPP--NHTVTTT 621
Db 697 SSTATNSASSETPCNSETOTSDGTSTMTVSSDSTTAGPSSTATNSASSETPCNSETOTSD 756
Qy 622 EYWSQSFATTTTIVTAPPG--TDTVIREPPNHTVTTTEYWSQSFATTTTITAPPGETDV 680
Db 757 GTFTMTVPNDSTTAGPSSTATNSASSETPCNSETOTSDGTSTMTVSSDSTTAGPSSTAT- 815
Qy 681 LIREPPNHTVTTTEYWSQSFATTTTIVTAPPGTDT--TAPGETDT--VLIREPPNHTVTTTEYWS 733
Db 816 --NSASSETPCNSETOTSDGTSTMTVPNDSTTAGPSSTATNSASSETPCNSETOTSDGTS 873
Qy 734 QSYATTTTIVTAPPGTDT--VIREPPNHTVTTTEYWSQSFATTTTIVTAPPGT----- 785
Db 874 TMTVPNDSTTAGPSSTATNSASSETPCNSETOTSDGTSTMTVPNDSTTAGPSSTATNSAS 933
Qy 786 -DTVIIYBSMSSSKISTSS--NDITSIIPFSRPHYVNSTT---SDLSTFESSMNTPTS 839
Db 934 SETPCNSETOTSDGTSTMTVPNDSTTAGPSSTATNSASSETPCNSETOTSDGTSTMTVPN 993
Qy 840 ISSDGMLLSSTTLVTESET--TTELICSDGKEC-----SRLSSSGIVTNDPSNES--S 889
Db 994 DSTTAGPSSTATNSASSETPCNSETOTSDGTSTMTVPNDSTTAGPSSTATNSASSETPCN 1053
Qy 890 IVTSVTPASTWS--DSLSTDGISATSDNVKSGVSTTETSVTTTQIOTPNPLSSSVTS 948
Db 1054 SETQTSOGTSTMTVSSDSTTGPCSSTATNSASCEPFFVSHKECKYDSTTAGPSSTATNS 1113
Qy 949 LPQLSSIPSVESE-----SKVFTSNGDNQSGTHDSQSSTSTEIE 988
Db 1114 AS--SETPCNSETOTSDGTSTMTVPNDSTTAGPSSTATNSASSETPCNSETOTSDGTFTM 1171
Qy 989 IVTTSSTKVLPPVSSNTDLTSEPTNTREOPTTLTSTNSITLITTSOPTGNGDNTSS 1048
Db 1172 TVSSDSTTAGPSSTATNSASSETPCNSETOTSDGTFTMTVSSDSTTAGPSSTATNSAS 1230
Qy 1049 TNP-----VPTVATSTLASAEEDNKSGHESASTSLKPSMGNSGLTTS 1093
Db 1231 ETPCNSETOTSDGTSTMTVPN--DSTAGPSSTATNSASSETPCNSETOTSDGTSTMTVS 1288
Qy 1094 TEIEATTSPTPEAPSVSSGTDVTEPTDREOPTTLTSTKINSSELVATQATNENG 1153
Db 1289 S--DSTTAGPSSTATNSASSETPCNSE--TQTSOGTSTMTVSSDSTTAGOSST--ATNSASS 1344
Qy 1154 KSP-----STDLT-----SSLTGTASATSANSELVTSVSGVTGGAVASA 1192
Db 1345 ETPCNSETOTSDGTSTMTVSSDSTTAGPSSTATNSASSETPCNSETOTSDGTSTMTVSSD 1404
Qy 1193 SNDQSHSTSVTNSIVSNTPQTLTSSQVTSSTNTF 1231
Db 1405 STTAGOSSTATNSAS--SETPCNSETOTSDGTSTMTF 1439
```

RESULT 15

S25345

probable membrane protein YCR089w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YCR1102

C:Species: Saccharomyces cerevisiae

C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000

C:Accession: S25345; S19504

R:Wilson, C.; Grisanti, P.; Frontali, L.

Yeast 8, 569-575, 1992

A:Title: The complete sequence of a 6146 bp fragment of Saccharomyces cerevisiae chro

A:Reference number: S25345; MUID:92397594; PMID:1523889

A:Accession: S25345

A:Molecule type: DNA
A:Residues: 1-1609 <WIL>
A:Cross-references: GB:X59720; EMBL:S43845; NID:g1907116; PIDN:CAA42254.1; PID:g1907227
R:Frontali, L.; Grisanti, P.
submitted to the Protein Sequence Database, March 1992
A:Reference number: S19504
A:Accession: S19504
A:Molecule type: DNA
A:Residues: 1-1609 <PRO>
A:Cross-references: EMBL:X59720; NID:g1907116; PID:e264634; PID:g1907227; MIPS:YCR089w
C:Genetics:
A:Gene: SGD:FIG2
A:Gene-references: SGD:S0000685; MIPS:YCR089w
A:Map position: 3R
C:Keywords: transmembrane protein
F:4-20/Domain: transmembrane #status predicted <TM1>
F:1592-1609/Domain: transmembrane #status predicted <TM2>

Query Match 8.4%; Score 547.5; DB 2; Length 1609;
Best Local Similarity 22.5%; Pred. No. 3e-16;
Matches 332; Conservative 232; Mismatches 503; Indels 411; Gaps 66;

QY 55 SLDGTANPCDTE-LNMPVFKYTTQSVDLTADGVKATQCFYSGEEFTFSLTCT 113
DB 165 SLFTSVNPSQSWTSENSE---KSSALSSTIDFTSSEISGSTSP-KSLESEFDTGTTSS 220
QY 114 VNDA-----LKSIRAF---GTVLPPIAFNVGGTGSST 143
DB 221 YSPSPSKNSNOTSLLSPLEPLSSSGDLILSTIQATNDQTSKIPIVLDATSSLPT 280
QY 144 DLEDSKCFRAGTNVTFN-----DGDKDISIDVEFEKSVDPSPAYL-----YASRVMP 191
DB 281 LRSSMAPTSGDSISHNFTSPPKTSGNYDV-----LTSNIDPSLEFTTSEYSTQLS 335
QY 192 SLNKVT-----TLFVAPQCENGT-SGTMGFSSNGDVAIDCSN 229
DB 336 SLNRASKSETVNFATSIATPFGTDSATSLIDPISVSGTASFGVIGISTAFSTQGN 395
QY 230 IHGITKG---LNDW---NPVS-----SEFSYTKTKTSNGIOI 263
DB 396 VPSTAGSSQYQDNSSSLPLSQTTWVINTNTNGSVTSTSPAYVATKTVQGVIT 455
QY 264 KY-----QNPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLOQSPFTLRWG 314
DB 456 EYVWCPPLQTKQAIGVSSISSVPQASFGSSILSSNSTLAAS--NNVPESTAGS 513
QY 315 YKNSDAGSGIVAVT-----RTVDSITAVTTLFPNPSVDKTKI-----EILQPIPT 364
DB 514 SQQDNSSSLPLSQTTWVINTNTQGSVTSTSPAYVST-ATKTVQGVITEYVWCP 572
QY 365 TTTTSYVGVTTSLYRKT-----APIGETATVVDVPYHTTTVTSEWGTITTTT 416
DB 573 TQTKQAIGSSITATQTSKPSSILTLGISLQLSDATFKGTETIN---THLMTESIS 629
QY 417 RINPT-----DSIDVWVQVLPNPTVSTTEYWSQSFATTTVTAPPQGTDTVIIREPP 470
DB 630 ITEPTVFSGTSDFYLCSEVNLAS-SLSYPNFSSEGSTATIT----- 673
QY 471 NHTVT---TTEYWSQSFATTTVTA-----PPGGTDSVIIREPPNPVTVTTEY 515
DB 674 NSTVTFGSTSKYPSVNSPTEASQHVSSVNSLDTFTNSTETIAVINSIHKTSNKDY 733
QY 516 WSQSFATTTVTAPPQGTDSVIIREPPNPVTVT-----TEY-W--SQSYATTTVTAP 565
DB 734 -----SLITTQLKTSKQTLVL-----STVITTVNGAATEYTWCPASSIATTSISY- 781
QY 566 PGCTDSVIIREPPNHVTTTEYWSQSFATTTVTAPPQGTDTVIIREPPNHV---TTTE 622
DB 782 -----KTLVLATEVCHSECTPVITSVATSSIPLLSSTSSVLSSTVSE 828
QY 623 YNSQSFATTTVTGPPSGTDTVIIREPPNPVTVTTEYWSQSFATTTITAPPGETDTVLI 682
DB 829 GAKNPAASEVTINTQVSATS-----EATSTSQ-----VSATSATASESTTSQ 874

Search completed: June 11, 2003, 17:14:06

Job time : 42 secs

QY 683 REPPNHTVTTEYWSQSYATTTVTAPPGETD-----TVLIREPNNHTVTTEYWSQSYAT 738
DB 875 VSTASETIST--LGTONFTTGSLLFPALPSTEMINTVVSRTK---LIISTEVCSHKKCV 929
QY 739 TTTVT-----APPGG-----TDTVIIREPPNPVT--TTEYWSQSFATTTVTAPP 783
DB 930 PTVITEWVTSKGTPSNGSHSSQTLQTEAVEVTLSHQVTMTSTMEVCNSICTPTVITSVQ- 988
QY 784 GTDTVILYISMSSSKISTSNIDTISIIPSFRPHYVNSTISDLSTESSMNTPTSISSD 843
DB 989 MRSTPPPYLTSTSSSLASTKKSSL-----EASSEMSTFSVTSQSLPLAFTCS 1037
QY 844 GMLLSSTTLVTESETT--TELICDCKECSRLSSSGIVTN--PDGNES-----SIV 891
DB 1038 EK--RSTTVSQMSNVILNTIMS-----SSNVISTNEKPSSTTSPYFNSSGYSLP 1087
QY 892 TSTVPTASTMSLSSTDGI-----SATSSDNVS 920
DB 1088 SSGTSPQSYLSLSTATTINGIKVTWTCPLAEKSTVAASSQSSRSVDRFVSSSKPSSLS 1147
QY 921 KSCGVSVTTSTVTIQ-----TTPNPLSSVT--SLTQLSSITPSVSESESKVTFTSNGD 972
DB 1148 QTSIQTLSTATTISGLKTVVTTWCPLTSKSLGATTOTS-----STAKVRIYS-AS 1199
QY 973 NQSGTHDSQSTSTEIEIVTTSSTKVLPPVSSNTOLTSE--PTNTRQPTTLS-----TTS 1026
DB 1200 SATSTSISLSTFSE-----SSGGLSKGVCSTECTQDVPTQSSSPASTLAYSPSVSTS 1255
QY 1027 NSTEDITTSQPTGNDNTSSTNPVPTVATSTLASAEEDNKGSGHESASTSLKPSMGE 1086
DB 1256 SSSSFSTTTASTL-----TSTHTSVPLLPSSSSSISASSPSTS---LLSTSLPSPAF 1305
QY 1087 NSGLTTSIEATTTSTPTAPSPVSSGTDVTEPTDTRQPTTLSTTKTNSEL----- 1141
DB 1306 SSTPLPTATAVSSSTFTIASSLP---LSSKSLSLSPVSSSILMSQFSSSSSSSLASLPS 1362
QY 1142 -----VATTOATNENGKSPST-----DLTSSLT-----TGTS 1169
DB 1363 LSLSPVTDIVSVLQPTTSIATLTCTDSQCOQEVSTICNGNCDVDTSTATPPSTVTD 1422
QY 1170 ASTSANSSELVTSQSVTG-----GAVASAND--QSHSTSVTNS-----NSIVS 1210
DB 1423 TCTGSECOQTSSSCDGYCKVSEYKSSATISACSGEGCQASATSELNSQYVTMTSVIT 1482
QY 1211 NTPQTTLSQOV--TSSSPSTNTFIASY---DGSGSII 1243
DB 1483 PSAITTSVEVHSTESTISITTVKVPVTTSSDNTNGELI 1520

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2004, 20:45:33 ; Search time 4785 Seconds
(without alignments)
10644.198 Million cell updates/sec

Title: US-09-715-876-7_COPY_52_1296
Perfect score: 1245
Sequence: 1 aagacaatcactggtgtttt.....cagtggtgtacaagttcca 1245

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_on.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_nam.*
- 37: em_htg_vrt.*
- 38: em_sv.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1245	100.0	3786	8	YSAALS1	L25902 Candida alb
2	1046.6	84.1	3813	8	AF068866	AF068866 Candida a
3	1037	83.3	4723	8	AF025429	AF025429 Candida a
4	1030.6	82.8	4044	8	AY227439	AY227439 Candida a
5	1029	82.7	4152	8	AY227440	AY227440 Candida a
6	949	76.2	3144	8	AY223551	AY223551 Candida a
7	949	76.2	4383	8	AF051313	AF051313 Candida a
8	947.4	76.1	3360	8	CAU87956	U87956 Candida alb
9	945.8	76.0	3468	8	AY223552	AY223552 Candida a
10	895	71.9	1404	8	CAIALS1	AF024580 Candida a
11	890.2	71.5	1404	8	CANALS2S1	AF024582 Candida a
12	733.4	58.9	1407	8	CANALS4S1	AF024584 Candida a
13	733	58.9	4569	8	AF272027	AF272027 Candida a
14	731.8	58.8	1007	8	AF202529	AF202529 Candida d
15	728.6	58.5	1407	8	CANALS42S1	AF024586 Candida a
16	713.8	57.3	1404	8	AF229898S1	AF229898 Candida a
17	713.8	57.3	5502	8	AY269422	AY269422 Candida a
18	713.8	57.3	5565	8	AY269423	AY269423 Candida a
19	574.6	46.2	4332	8	AF075293	AF075293 Candida a
20	573	46.0	4101	8	AY225310	AY225310 Candida a
21	507	40.7	995	8	AF202530	AF202530 Candida d
22	463.6	37.2	1013	8	AF201685	AF201685 Candida d
23	393.6	31.6	6897	8	AF201684	AF201684 Candida a
24	392	31.5	1413	8	AY296650	AY296650 Candida a
25	390.6	31.4	1299	8	AF075294	AF075294 Candida a
26	362.2	29.1	1058	8	AF201686	AF201686 Candida t
27	141.6	11.4	377	8	AF211866	AF211866 Candida t
28	132	10.6	380	8	AF211865	AF211865 Candida t
29	125.4	10.1	1400	8	AB002099	AB002099 Candida t
30	112.2	9.0	157	8	AF035757	AF035757 Candida a
31	105	8.4	170221	10	AC127421	AC127421 Mus muscu
32	103.4	8.3	216959	10	AC116998	AC116998 Mus muscu
33	90.2	7.2	252220	2	AC122672	AC122672 Rattus no
34	88.2	7.1	170669	2	BX322580	BX322580 Danio rer
35	87.8	7.1	146907	2	AC141031	AC141031 Rattus no
36	87.8	7.1	158615	2	AC117835	AC117835 Rattus no
37	86.8	7.0	203271	2	BX322566	BX322566 Danio rer
38	85.6	6.9	176322	2	AL935062	AL935062 Danio rer
39	85	6.8	35793	5	AY016024	AY016024 Takifugu
40	83.8	6.7	266451	2	AC098493	AC098493 Rattus no
41	82.4	6.6	100	6	A43858	A43858 Sequence 32
42	82.4	6.6	100	6	A43860	A43860 Sequence 34
43	82.4	6.6	100	6	I17435	I17435 Sequence 32
44	82.4	6.6	100	6	I17437	I17437 Sequence 34
45	82	6.6	158266	2	BX323881	BX323881 Danio rer

ALIGNMENTS

RESULT 1
YSAALS1
LOCUS
DEFINITION
Candida albicans agglutinin-like sequence (ALS1) gene, complete cds.
3786 bp DNA linear PLN 03-MAY-2000
ACCESSION
L25902
VERSION
L25902.1
KEYWORDS
GI:704426
SOURCE
Candida albicans
ORGANISM
Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
1 (bases 1 to 3786)
AUTHORS
Hoyer,L.L., Scherer,S., Shatzman,A.R. and Livi,G.P.
TITLE
Candida albicans ALS1: domains related to a Saccharomyces

Saccharomycetales; mitosporic Saccharomycetales; Candida.

```

REFERENCE 1 (bases 1 to 3813)
AUTHORS Hoyer,L.L. and Hecht,J.E.
TITLE The ALS5 gene of Candida albicans and analysis of the Als5p
N-terminal domain
JOURNAL Yeast 18 (1), 49-60 (2001)
MEDLINE 21064501
PUBMED 11124701
REFERENCE 2 (bases 1 to 3813)
AUTHORS Hoyer,L.L., Ho,M. and Hecht,J.E.
TITLE The ALS5, ALS6 and ALS7 genes of Candida albicans
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 3813)
AUTHORS Hoyer,L.L., Ho,M. and Hecht,J.E.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1998) Veterinary Pathobiology, University of
Illinois at Urbana-Champaign, 2001 S. Lincoln Avenue, Urbana, IL
61802, USA

FEATURES
     source              Location/Qualifiers
     1..3813
        /organism="Candida albicans"
        /mol_type="genomic DNA"
        /strain="1161"
        /db_xref="taxon:5476"
        /chromosome="3"
        /note="SfiI fragment p"
     gene
     1..3813
        /gene="ALS5"
        /allele="ALS5-1"
     CDS
     1..3813
        /gene="ALS5"
        /codon_start=1
        /transl_table=12
        /product="agglutinin-like protein"
        /protein_id="AAD32849.1"
        /db_xref="GI:4903269"
        /translation="MIQFTLLFLYLSPATAKAITGIFNSIDSLTWSNAGNYAFKPGP
        YPTWNAVGLWSDTGANPGDFTILNMPGVKFTASQKSDVLTADGVKYATCOFYSGE
        EFTFTSTLTCDNALKSIKAFGTWLPPIAFNVGSGTSDLEDKSGKTAGINTVTF
        NDGSKLAIANFESTKVDKSGYLTSPKMPSLNKIATLYVAPQCEGVTSQMGFST
        SYGDVAIDCSNVHIGISGVNDVNHPTVSSEFSYKSGSCSSFGISITYQNVAGRPFI
        DAYTSNDQVLSKNDYTCDDYDWHAPFTLKWGTGKNSDAGSNGIVAVITRTV
        TDSPTAVTLPNPSVDKTKTIEILOPPTTITTSYVGVTSYKTAPIGETATLI
        VDVPVHTTTLTTSKMTGKTITMTTRNPTDSIDTIVVQVPLNPTTTTQFWSBFTS
        TTTINKPEGDSVIVKPHNPTVTTFWSEYATTTITPLGTDQSVIHDPLEE
        SSSPTAISDSNDSISQAQSSSVQSSIVGLSSSDIPLSSDMPSSSGTLESSE
        SSTVSDSSSSSELSTFSSSESYSSISDITNFWSSSDSLESTITWSSSIDAQ
        SDDPHTTIAGPDRSSSSIASIVISDLVSLTSDPTSSFDSSSLNSDSSSPFSD
        ESDISASSFSFLVAPSPFLSSSSLSLIYHYVNSTYTHASESSSVAGSPSVASES
        ANDDHTLSESDTTSIIIGTDSSTVTFRRDNGDCIVTGITSSSIDSEQTSDVTTTS
        SFVASSTPFSAQSDITNPNIDSSQTSASSTSKSSVSVDVTVNSILLSESTLSDD
        GTSDTSISSTSDTGNINAGSHSHKSTASIKESIQTGKTVLSSSYLSTKLSSDSDI
        TIELTTLTIEDNEPNTFTSPSSHEISLSDNSVLKQVDRSTIKTPTDVT
        VSSLSVHSTAEATLGENSFENAVSTPLNTATSLRSTSSSNHATESSTGVKSEASV
        RAIPGPSTNRLSLYSTEAEGLTYANSQTNLITISOVAAPTDSVLITENLVV
        STFDNSAANDQPSKTSIESIMNPDSNTEQNGFIATLSQAOVPSSSIHSELIST
        TTKATDASMGDSASNSQPTLLIQVATSSYNQPLITVYAGSSSATKHPFWLLKFI
        SVALPFFL"
BASE COUNT      1172 a      819 c      641 g      1181 t
ORIGIN
Query Match      84.1%; Score 1046.6; DB 8; Length 3813;
Best Local Similarity 90.0%; Pred. No. 1.3e-182;
Matches 1121; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY      1  AAGCAATCAGTGGTGTGTTTTCATAGTATTTTAAATCAATTAACCTTGGTCCAATGCTGCTAAT 60
Db      1  AAGCGATCAGTGGTATTTTCAATAGTATTTGACTCATTAACTTGGTCCAATGCTGCAAT 111
QY      61  TATGCTTTCAAGGCGCAGGATACCACACTTGAATGCTGTTTGGGTGGTCTTAGAT 120
Db      112  TACGCTTTCAAGGCGCAGGATACCACACTTGAATGCTGTTTGGGTGGTCTTAGAT 171
QY      121  GGTACCAAGTCCCAATCCAGGGGATACATTCACATTTGAATATGCCATGTGTGTTAAATAT 180
Db      172  GGTACCAAGTCCCAATCCAGGGGATACATTCATATTTAAACATGCCATGTGTGTTAAATAT 231
QY      181  ACTACTTTCACAAAATCTGTGTGTTTAACTCCGATGGTGTGTTAAATATGCTACTTGTCAA 240
Db      232  ACTGCTTCCCAAAAATCTGTGTGTTTAACTCCGATGGTGTGTTAAATATGCTACTTGTCAA 291
QY      241  TTTTATCTGCTGAAGAATTCACAACTTTTCTACATTAACATGCTACTGTGAACGCGCT 300
Db      292  TTTTATCTGCTGAAGAATTCACAACTTTTCTACATTAACATGCTACTGTGAACGCGCT 351
QY      301  TTGAATCATCCATTAAAGGCAATTTGGTACAGTTACTTTACCAATTTGCAATTTCAATTTGCT 360
Db      352  TTGAATCATCCATTAAAGGCAATTTGGTACAGTTACTTTACCAATTTGCAATTTCAATTTGCT 411
QY      361  GGAACAGGTTTCATCAACTGATTTGGAGATTTCTAAATGTTTTTACTGCTGGTACCAATACA 420
Db      412  GGAACAGGTTTCATCAACTGATTTGGAGATTTCTAAATGTTTTTACTGCTGGTATCAATAG 471
QY      421  GTCACATTTAATGATGGTGATAAAGATATCTCAATTTGATTTGAGTTTGAAGTCAACC 480
Db      472  GTAACATTTAATGATGGTGATAAAGATATCTCAATTTGATTTGAGTTTGAAGTCAACC 531
QY      481  GTTGATCCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGTTCACA 540
Db      532  GTTGATCCAAGTGGGTAATTTGACTACTTTCCAGATTTTATGCCAGTCTCAATAAATTTGCT 591
QY      541  ACTCTTTTGGGACCAACATGTAAGATTTGGTACATCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db      592  ACTCTTTTGGGACCAACATGTAAGATTTGGTACATCTGCTGCTGCTGCTGCTGCTGCTGCT 651
QY      601  AGTAGTACGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db      652  AGTAGTACGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 711
QY      661  TTAATGATTTGGAATTTATCCGTTTTCATCTGAATCATTTAGTTACATCAATAAATCTTGTACA 720
Db      712  GTAAATGATTTGGAATTTATCCGTTTTCATCTGAATCATTTAGTTACATCAATAAATCTTGTACA 771
QY      721  TCTAATGGAATTCAGATTTAAATATCAAAATGTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db      772  TCTTTTGGTATCTCTATCACATATCAAAATGTTCTCGCGGTATTCGTCCTCAATTTTATTTGAC 831
QY      781  GCTTATATTTCTGCTACAGATTTAAACCAATATCTTTTAGCATATACCAATGATTAATCTACT 840
Db      832  GCTTATATTTCTGCTACAGATTTAAACCAATATCTTTTAGCATATACCAATGATTAATCTACT 891
QY      841  TGTGCTGCGAGTCTGCTCTCAAAAGTAAACCTTTTCACTTTAAGATGAGCTGGATACCAAGAT 900
Db      892  TGTGTTGATGATTTATGGCAACATGCACCTTTCACTTTTAAATGAGCTGGATATAAGAT 951
QY      901  AGTGATCGCGGATCTAAGCGTATTTGCTATGTTGCTACAACTAGAACAGTTACAGACAGT 960
Db      952  AGTGATCGCGGATCTAAGCGTATTTGCTATGTTGCTACAACTAGAACAGTTACAGACAGT 1011
QY      961  ACCAGTGTGCTACCTACTTTTACCAATTTCAATCCAGTGTGTTAATAACCAAAACCAATCGAA 1020
Db      1012  ACCAGTGTGCTACCTACTTTTACCAATTTCAATCCAGTGTGTTAATAACCAAAACCAATCGAA 1071
QY      1021  ATTTTGGCAACCTATTTCCAAACCACTACCATCAACTTCAATTTGTTGGTGTGCTACTTCC 1080
Db      1072  ATTTTGGCAACCTATTTCCAAACCACTACCATCAACTTCAATTTGTTGGTGTGCTACTTCC 1131
QY      1081  TATCTGACTAAGACGTGACCAATTTGGTGAACAGCTACTGTTATTTGTTGATGTGCTAT 1140
Db      1132  TATCTGACTAAGACGTGACCAATTTGGTGAACAGCTACTTCTTTATTTGTTGATGTGCTAT 1191
QY      1141  CATACTACCAACATGTTTACCAGTCAATGGAGGAGCAATCACTACCAACCAACTCGT 1200
Db      1192  CATACTACCAACATGTTTACCAGTCAATGGAGGAGCAATTTACGATGACTACCAACTCGT 1251

```

QY 1201 ACCAATCCAACTGATTAATGACACAGTGGTGGTACAAAGTTCCA 1245
 |||||
 Db 1252 ACCAATCCAACTGATTAATGACACAGTGGTGGTACAAAGTTCCA 1296

RESULT 3
 AF025429 4723 bp DNA linear PLN 16-DEC-1997
 LOCUS
 DEFINITION Candida albicans agglutinin-like adhesin (ALAI) gene, complete cds.
 ACCESSION AF025429
 VERSION AF025429.1 GI:2522218
 KEYWORDS
 SOURCE
 ORGANISM
 Candida albicans
 Candida albicans
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporic Saccharomycetales; Candida.
 1 (bases 1 to 4723)
 Gaur, N.K. and Klotz, S.A.
 Expression, cloning, and characterization of a Candida albicans
 gene, ALAI, that confers adherence properties upon Saccharomyces
 cerevisiae for extracellular matrix proteins
 Infect. Immun. 65 (12), 5289-5294 (1997)
 JOURNAL
 MEDLINE 98053977
 PUBMED 9393828
 REFERENCE
 2 (bases 1 to 4723)
 Gaur, N.K. and Klotz, S.A.
 Direct Submission
 TITLE
 Submitted (15-SEP-1997) Research Service, VA Medical Center, 4801
 Linwood Blvd., Kansas City, MO 64128, USA
 JOURNAL
 Location/Qualifiers
 1. .4723
 /organism="Candida albicans"
 /mol_type="genomic DNA"
 /db_xref="taxon:5476"
 327. .4586
 /gene="ALAI"
 327. .4586
 /gene="ALAI"
 /function="cell adhesion protein"
 /codon_start=1
 /transl_table=12
 /product="agglutinin-like adhesin"
 /protein_id="AAB88883.1"
 /db_xref="GI:2522219"
 /translation="MIQFETLLFLXLSFATAKATGIPNSIDSLTWSNAGNVAFKGPG
 YPTWNAVLGSDGTSANFGDTFILNMPVFKPTASQKSDVLTAGVKYATCQFVSGE
 EFTFSLSKCTVNNRSLKALGTVLPINRVGGTSGVSDLEDSKCFAGTNTVTF
 NDGKKLSIAVNFKSTVDQSGYLITRPMPLINKIATLYVAPQENGVTGTMGFST
 SYGDVAIDCSNVHIGISKGVDNWHPTVSESFSYKSCSFGISITYQNPAGYRPF
 DAYISPDNNVOLSKNDYTCVDDYVQHAPFLKWTGYNKSDAGSNGIVIVATRTV
 TDSATVTLPLNPVDKTKTIELLOPIPTTITTSYGVVTSYSTKAPIGETATVI
 VDVPYHTITVSEWGTGTTTTTRTNPDSIDTVVQVPSNPNTTTFQWSEFTS
 TTTITNSLGTDSVIVREPHNPVTTFEWSFATETITKPGTDSVIVREPHNP
 TTTTTFEWSYATETITNGEGSDSVIVREPHNPVTTFEWSYATETITKPK
 EGDVSVIVKPNPTVTTFEWSYATETITNGEGSDSVIVREPHNPVTTFEWF
 SESYATETITPLGTDSIVIHDLPESSSTTAIESDSNTSSAQSSESSVEOSS
 IYGLSSSDIPLSSDMPSSSTGLSSESSTVSSYSDSSSIESSTLSSDRCSSSI
 SDTISFWSDDLESITWSSSIDAQSHLVQSVNSISIGSELSSSESSESTFA
 TDALVSSDASSLSSTSYPSSTISSDDPHTIAGSDSLISFITSTVEISDS
 VLSLTPASFPSSSGLNSDSSPSDQSDILTSSSFSTLVVPSFSLSSSSLSTLY
 PHVNTSTVHASESSSVASPMASESANDDTYLTSESTDTTSSGTDSSTVFCR
 DNGDCIVTGMPSSSIISQTSQDVTTTSSFVASSPTSAEQSITDNPNTDSSOTASS
 STKSSVSQDVTVVNSILLSETSLSSDSTSDSTISSTNSDTGGINAGSSHTSPAS
 IKESSIKGTGVNSLSYLSKJSSDITIELLITELITIELITIELITIEDNPFTSTPS
 SHSEIFSPNSVLSKQVDRESTITKTPDVTVDVTVSSLSVHSTEASTATIGENSPNVA
 STPSNTATSLRSTSSSNHATSSGTGVKSEAGAEIIPSPSTDNRSLYSTBEAKGIT
 YANSGSTNLIITESQVAAPTDSVLIENPVVTSPTDNNSSAAVDOPSKTKSIESIM
 NPDSTWETNGFTATLSQAQVPSSTSHSELISITTTAKTTDASNMNGSAAASNQPTLLI
 QQVATSSYNQPLITTYAGSSATKHPKSLKFLISVALPFFL"
 BASE COUNT 1441 a 1013 c 803 g 1466 t
 ORIGIN
 Query Match 83.3%; Score 1037; DB 8; Length 4723;

Best Local Similarity 89.6%; Pred. No. 7e-181;
 Matches 1115; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 1 AAGACAATCACTGGTGTGTTTGTATAGTATTTTAAATTCATTAACTTGCTGCTCAATGCTGCTAAT 60
 |||||
 Db 378 AAGCGATCACTGGTATTTTTCATATAGTATTTGACTCATTAACCTTGGTCCATGCTGGCAAT 437

QY 61 TATGCTTTTCAAGAGGCCAGGATACCCAACTTGGAAATGCTGTTTGGTGGTGGTCTTGTAGAT 120
 |||||
 Db 438 TAGCGTTTCAAGAGGCCAGGATACCCAACTTGGAAATGCTGTTTGGTGGTGGTCAATAGAT 497

QY 121 GGTACAGTGGCAATCCAGGGGATACATTCACATTTGAATATATGCCATGCTGTTGTTAAATAT 180
 |||||
 Db 498 GGTACAGTGGCAATCCAGGGGATACATTCATATTTAAACATGCCATGCTGTTGTTAAATAT 557

QY 181 ACTACTTCCAAAACATCTGTTGATTTAACTGCGAATGGTGTAAATATGCTACTTGTCTCAA 240
 |||||
 Db 558 ACTGCTTCCAAAACATCTGTTGATTTGACTGCGAATGGTGTAAATATGCTACTTGTCTCAA 617

QY 241 TTTTATTTCTGGTGAAGAAATTCACAACTTTTCTCATATTAACATGCTACTGTGAACACGCT 300
 |||||
 Db 618 TTTTATTTCTGGTGAAGAGTTTACAACTTTTCTCATTTAAATGTACAGTGAACAATAT 677

QY 301 TTGAAATCATTCATTAAGGCAATTTGGTACAGTTACTTTACCAATTTGCAATTTCAATGTTGT 360
 |||||
 Db 678 TTGAGATCATCTATTAAAGCTTTGGGTACGGTTACTTTACCAATTTGCAATTTCAATGTTGT 737

QY 361 GGAACAGGTTTCATCAACTGATTTGGAAGATCTTAATGTTTCTGCTGCTGTTGACCAATACA 420
 |||||
 Db 738 GGAACAGGTTTCATCACTGATTTGGAAGATCTTAATGTTTCTGCTGCTGTTGACCAATACG 797

QY 421 GTCACTATTAATGATGGTGATAAAGATATCTCAATTTGATTTGAGTTTGAAAGTCAACC 480
 |||||
 Db 798 GTAACTATTAATGATGGTGATAAAGATCTCAATTTGCTGTTTAAATTTGAAAGTCAACA 857

QY 481 GTTGATCAAGTGCAATATTGTTGATGCTTCCAGAGTTATGCGCAAGTCTCAATTAAGTCAACA 540
 |||||
 Db 858 GTTGATCAAGTGCGGTATTGACTACTTCCAGATTTATGCGCGAGTCTCAATATAAATTTGCT 917

QY 541 ACTCTTTTGTGGGACCAACATGTCAAAATGGTTACACATCTGTTGACCAATGGGTTCTCC 600
 |||||
 Db 918 ACTCTTTTGTGGGACCAACATGTCAAAATGGTTACACATCTGTTGACCAATGGGTTCTCC 977

QY 601 AGTAGTAACGGTGAAGTGTCTATTGATTTGCTCAAAATATTTCATATTGGTATCAAAAAGA 660
 |||||
 Db 978 AGTAGTATGGGATGTTGCTATTGCTCAATGATGATATATTTGATTTTCAAAAAGA 1037

QY 661 TTTAAATGATTTGAAATTTACCGGTTTCACTGAAATCATTTAGTTTACATTAATACTGTACA 720
 |||||
 Db 1038 GTAAATGATTTGGAATCATCCAGTACGTTCACTGAATCATTTAGTTTACACTAAAAAGTTGTTCA 1097

QY 721 TCTAATGGAAATTCAGATTAATAATCAAAATGTAACCTGCTGCTGTTTATGCTCATTTATTGAT 780
 |||||
 Db 1098 TCTTTTGGTATCTCTATCACATATCAAAATGTTCTTCCGGTATTCGTCATTTATTGAC 1157

QY 781 GCTTATATTTCTGCTACAGATGTTAAACCAATATCTTTAGCATATATCAATGATATATACT 840
 |||||
 Db 1158 GCTTATATTTCTCCCTCAGATTAATTAACAGTATCAATTTGCTGATATAAATGATATACT 1217

QY 841 TGTGCTGGCAGTGGTCTGCAAAAGTAAACCTTTTCACTTTTAAGATGGAAGTGGATCAAGAAT 900
 |||||
 Db 1218 TGTGTTGATGATTAATTTGGCAGCATGCACCTTTTCACTTTAAAAATGGAGTGGATATAAGAAT 1277

QY 901 AGTGATGCCGATCTAAACGGTATTTGCTATTTGCTGCTACAACTAGAACAGTTCACAGACAGT 960
 |||||
 Db 1278 AGTGATGCCGATCTAAACGGTATTTGCTATTTGCTGCTACAACTAGAACAGTTCACAGACAGT 1337

QY 961 ACCACTGCTGTCACTACTTTTACATTTCAATTTCAATTTCAATTTGATTAATAAACAATTCGAA 1020
 |||||
 Db 1338 ACCACTGCTGTCACTACTTTTACATTTCAATTTCAATTTCAATTTGATTAATAAACAATTCGAA 1397

QY 1021 ATTTTGCACACTATTTCACCAACCACTACCATCAAACTTCATATATTTGCTGCTGCTACTTCC 1080
 |||||

```

Db      1398 ATTTCGAACCTATTCCACCACTACCATCAACATTCATATATGTGTGGTGTGACTACTTCC 1457
Qy      1081 TATCTGACTAAGACTGCACCAATTTGGTGAACAGCTACTGTTATTGTTGATGTGCCATAT 1140
Db      1458 TATCTGACTAAGACTGCACCAATTTGGTGAACAGCTACTGTTATTGTTGATGTGCCATAT 1517
Qy      1141 CATACTACCAACTGTTTACAGTGAATGGACAGGACCAATCACTACCAACCACTCGT 1200
Db      1518 CATACTACCAACTGTTTACAGTGAATGGACAGGACCAATCACTACCAACCACTCGT 1577
Qy      1201 ACCAATCCAAGTATTCAATTCGACACAGTGGTGGTACAAAGTTCCA 1245
Db      1578 ACAATCCAAGTATTCAATTCGACACAGTGGTGGTACAAAGTTCCA 1622

RESULT 4
LOCUS   AY227439
DEFINITION Candida albicans cell-surface adhesin (ALS5) gene, ALS5-small allele, complete cds.
ACCESSION AY227439
VERSION   2
KEYWORDS  AY227439.1 GI:29373078
SOURCE  Candida albicans
ORGANISM Candida albicans
REFERENCE 1 (bases 1 to 4044)
AUTHORS Zhao,X., Nuesen,J.A. and Hoyer,L.L.
TITLE Analysis of Candida albicans ALS5p function
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4044)
AUTHORS Zhao,X., Nuesen,J.A. and Hoyer,L.L.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2003) Veterinary Pathobiology, University of Illinois-Urbana, 2522 VMSB, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA

FEATURES
source
    1..4044
        /organism="Candida albicans"
        /mol_type="genomic DNA"
        /strain="SC5314"
        /db_xref="taxon:5476"
        /chromosome="6"
    <1..4044
        /gene="ALS5"
        /allele="small"
    <1..4044
        /gene="ALS5"
        /product="agglutinin-like protein"
    1..4044
        /gene="ALS5"
        /note="ALS5p; cell-surface adhesin"
        /codon_start=1
        /transl_table=12
        /product="agglutinin-like protein"
        /protein_id="AA072528.1"
        /db_xref="GI:29373079"
        /translation="MLQPTLLFLYLSPATAKAITGIFNSIDSLTWSNAGNVAFKPGP
        YPTWAVLGSMDLNNRSGIKALGVTLIPAFNVGSGTSSVDLEKCFAGTANTVTF
        BEPTFSSKGTVDNNSIKALGVTLIPAFNVGSGTSSVDLEKCFAGTANTVTF
        NDGKSLGVNFKSTVDOSGLTTSRPMPLKIALTLVVAPOCEGVSTGTMGFT
        SYGDVAIDCSNVHIGISKGVNDMHPVTSESFSYTKSCSGISITXQNVAPGYRPI
        DAYISPDNNQYQSKYKDYTCDDYQWHPFLRTWGYKNSDAGSNGIVIVATRTV
        TDSTVATLTPNPSVDKTKTLEILOPIPTTITTSYVGVITTSYTKTAPIGETATVI
        VDVPYHTTSTVTKGTTTITTTTNTPTSDIDTVVQVPLENPTTTTQFMSEFST
        TTTTINSGTDSVLRPNPTVTTFWSSSYATITINGPBGTDVIVREBPHN
        TTTTFWSSYATITVNTKPGTDSVIVKPNPTVTTFWSSYATITITTP
        LCTDSIVLHDLPESSSTTAIESDSNLSAQSSESSVQSSIVGLSSSDPLSS
        DNPSSSGLTLESSESTVSSYDSDSSESELSTFSRSGYSSSIDTNFWDSSDL
        ESTIWSSSIDAQSSQVQSVNSISTQETTSRSGSESNSTSVTDILVSSDASILN
        SDISYPSSTLSLDLPHFTIAGPDRSSSSIASTVEISDLSVLSLSDPTSSPDS
        SSLNSDSSSSPDESISASSSFSTLVAPSPSLSSSSLSLIYPHYVNSTTYHASES

```

```

ESSSVASPSMASSEANDDTHLSESDTSSIGTDSSTVTFCCRRNGDGCIVTGIPSS
SIDSEQSDVTTSFVSAQSTIDNPNDSSQTSASSKSSVSQSDTVN
SILLSETSLSSDSDSTGTSNTSDGTGNINAGSSTSTASIKESIQKTGVTL
SYLSTKLSSTSDITIELITELITELITEDNEPNTFTSPSSHSEIFSSDNLVLS
KQVDESTIKTSPPTDVTVSSLSVHSEASTATGENSPSNVASTPSNIATSLBSTS
SSNHATSESGTVKSEASAEIIPSPPTSTDRLSYSTBEAKGITVANGSGSTNNLINES
QVAAPDTSVLIEPNFVSTFDNNSAAVDGSPKTSIEESIMPDSTNETNNGFIA
TLSQAPVSSSIHSELISTTTAKTDDASMDGSAASNOPTTLLIOQVATSSYNQPLIT
TYAGSSATKHPKSWLLKPIVALPFL"
BASE COUNT      1242 a      881 c      685 g      1236 t
ORIGIN
Query Match      82.8%; Score 1030.6; DB 8; Length 4044;
Best Local Similarity 89.2%; Pred. No. 1.1e-179;
Matches 1111; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
Qy      1 AAGACAATCAGCTGGTGTGTTTATGATAGTATTATTCATTAACCTGGTCCAATCGCTAAT 60
Db      52 AAGCGGATCAGCTGGTATTTTCAATAGTATTTGACTCATTAACCTGGTCCAATCGCTAAT 111
Qy      61 TATGCTTTCAAGGCGCAGGATACCCAACTTGGAAATGCTGTTTGGGTTGGTCTTAGAT 120
Db      112 TACGCTTTCAAGGACAGGATACCCAACTTGGAAATGCTGTTGTTGGTGGTCAATAGAT 171
Qy      121 GTTACAGTGCCAATCCAGGGGATACATTCACATTAATGAATATGCCATGTGTGTTAAATAT 180
Db      172 GTTACAGTGCCAATCCAGGGGATACATTCATATTAACATGCCATGTGTGTTAAATATC 231
Qy      181 ACTACTTCACAAACATCTGTTGATTTAACTGCCGATGGTGTAAATATGCTACTTGTCAA 240
Db      232 ACTGCTTCCCAAAAATCTGTTGATTTGACTGCCGATGGTGTAAATATGCTACTTGTCAA 291
Qy      241 TTTTATCTGGTGAAGAATTCACAACTTTTCTACATTAACATGTACTGTGAACGACGCT 300
Db      292 TTTTATCTGGTGAAGATTTACAACCTTTTCTCATTAATAATGTACAGTGAACATAT 351
Qy      301 TTGAATCATCCATTAAGGCATTTGGTACAGTTACTTTTACCAATTCATTAAGTGTGT 360
Db      352 TTGAGATCATCTATTAAAGCTTTGGGTACGGTACTTTTACCAATTTGCAATTTGCT 411
Qy      361 GGAACAGGTTTCATCAACTGATTTGGAAGATTTCAATGTTTACTGCTGGTACCATAACA 420
Db      412 GGAACAGGTTTCATCAAGTTGATTTGGAAGATTTCAATGTTTACTGCTGGTACCATAACG 471
Qy      421 GTCACATTTAATGATGGTGAATAAGATATCTCAATTTGATTTGAGTTTGAAAGTCAACC 480
Db      472 GTAACATTTAATGATGGCAGTAAAGGCTCTCAATTTGCTGTTTAAATTTGAAAGTCAACA 531
Qy      481 GTTGATCCAAGTGCAATTTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATTAAGTCA 540
Db      532 GTTGATCAAAGTGGGTATTTGACTACTTCCAGATTTATGCCGAGTCTCAATTAATTTGCT 591
Qy      541 ACTCTTTTGGGACCAACATTTGGAATGGTTACACATCTGGTACAAATGGGGTTCTCC 600
Db      592 ACTCTTTTATGGGACCAACATTTGGAAGCGTTACACATCTGGTACAAATGGGGTTCTCC 651
Qy      601 AGTACTAAGCGTGAGCTGCTATTGCTATTCATTTGCTCAAAATATTCATATTGATATCA 660
Db      652 ACTAGTTATGGGATGTTGCTATTGACTGTTTCAAAATGATACATTTGGTATTTCAAAAGA 711
Qy      661 TTAATGATTTGGAATTTATCCGTTTTCATCTGAATCATTTAGTTACATAAAGCTGTACA 720
Db      712 GTAATGATTTGGAATCATCCAGTTACGCTTGAATCATTTAGTTACATAAAGCTGTACA 771
Qy      721 TCTAATGGAATTCAGATTAATATCAAAATGTATCTGCTGGTATTCGTCATTATTGAT 780
Db      772 TCTTTTGGTATCTCTATCACATATCAAAATGTTCCTGCGGTTATTCGTCATTATTGAC 831
Qy      781 GCTTATATTTCTGCTACAGATTTAAACCAATATCTTTAGCATATACCAATGATATAT 840
Db      832 GCTTATATTTCTTCCCTCAGATAAATACCAAGTATCAATTTGCTGATAAAGATGACTAT 891
Qy      841 TGTGCTGCAGTCTGCTCCTCAAAAGTAAACCTTTTCACTTTTAAAGATGACTGATACAAGA 900

```

[illegible][illegible]

FEATURES

source

gene

mRNA

CDS

BASE COUNT ORIGIN	1277 a	904 c	701 g	1270 t	LF _{FFFL} ^W	
Query Match	82.7%	Score 1029;	DB 8;	Length 4152;		
Best Local Similarity	89.2%;	Pred. No. 2.1e-179;				
Matches 1110; Conservative	0;	Mismatches 135;	Indels	0; Gaps	0;	
QY	1	AAGCAATCACTGGTGT	TTTTTGATAGT	TTTAAATTCATTA	ACTTGGTCCAATGCTGCTAAAT	60
DB	52	AAAGCGATCACTGGTA	TTTTCAATAGATATTGA	CTCATTAACTTGGTCCAATGCTGGCAAT	111	
QY	61	TATGCTTTCAAAGGCCAGGAT	ACCCAACTTGGAAATGCT	TTTTTGGGTGGTCTCTTAGAT	120	
DB	112	TACGCTTTCAAAGCACAGGAT	ACCCAACTTGGAAATGCT	GTGTTGGGTGGTCAATAGAT	171	
QY	121	GGTACCAAGTCCCAATCCAGGGGAT	ACATTCACATTTGAATATG	CCATGTGTGTTTAAATAT	180	
DB	172	GGTACCAAGTCCCAATCCAGGGGAT	ACATTCATATTAACATG	CCATGTGTGTTTAAATAT	231	
QY	181	ACTACTTTCCAAACATCTGTTGAT	TTTAACTGCCGATGGTCT	TAAATATGCTACTTTGTCAA	240	
DB	232	ACTGCTTTCCAAAAATCTGTTGAT	TTTGAATGCTGCGGATGGT	TTAAATGCTACTTTGTCAA	291	
QY	241	TTTTTAATCTGGTGAAGAAAT	TCACAACTTTTCTACATTA	ACATGTACTGTGAACGACGCT	300	
DB	292	TTTTTAATCTGGTGAAGAGTTT	ACACTTTTCTTCATTA	AAATGTACAGTGAACAATAAT	351	
QY	301	TTGAAATCATCCATTAAGGCAT	TTGGTACAGTTACTTTAC	CAATTTGCATTTCAATTTGGT	360	
DB	352	TTGAGATCATCTAATTAAGGCT	TTTGGGTACGGTTACTTTT	ACCAATGGCATTTCAATTTGGT	411	
QY	361	GGAAACAGGTTCACTCAATG	ATTTGGGAAGATTCCTAAAT	GTTTACTGCTGGTACCAATACA	420	
DB	412	GGAAACAGGTCATCAGTTGAT	TTTGGGAAGATTCCTAAAT	GTTTACTGCTGGTACCAATACG	471	
QY	421	GTCACTTTAATGATGGTGAT	ATAAGATATCTCAATTTGAT	GTTGAGTTTGAAGTCAACC	480	
DB	472	GTAACATTTAATGATGGCAGT	ATAAAGCTCTCAATTTGCT	GTTAAATTTGAAGAGTCAACA	531	
QY	481	GTTGATCCAAAGTGCAAT	TTTGTATGCTTCCAGAGT	ATTGCCAAGTCTCAATAAGGT	540	
DB	532	GTTGATCAAAAGTGGGTAT	TTTGAATTTCCAGATTTTAT	GCGAGTCTCAATNAATTTGCT	591	
QY	541	ACTCTTTTGTGGCACCAATG	CTGAAATATGGTTTACATCT	GGTACAAATGGGGTTCTCC	600	
DB	592	ACTCTTTATGTGGCACCAATG	CTGAAATATGGTTTACATCT	GGTACAAATGGGGTTCTCC	651	
QY	601	AGTAGTAAAGGTGACGTTG	CTCTAATTTGCTCAAAATAT	TTCAATTTGGTATCACAAAGGA	660	
DB	652	ACTAGTTATGGGGATGTTG	CTAATTTGCTCAAAATG	TACATATGGTATTTCAAAGGA	711	

QY 541 ACTCTTTTGTGGCACCACCAATGTGAAAATGGTTACACATCTGGTACAAATGGGTTCTCC 600
Db |||||
QY 592 ACTCTTTTGTGGCACCACCAATGTGCAATGGTTACACATCTGGTACAAATGGGTTCTCGT 651
Db |||||
QY 601 AGTAGTAACGGTGACGTTCTGCTATGATGCTCAAAATATTCATATTTGGTATCAAAAAGGA 660
Db |||||
QY 652 AACACTTATGGTGATGTTCAAAATGAGCTGTTCAAAATATTCATATTTGGTATTAACAAAGGA 711
Db |||||
QY 661 TTAATGATGGAATATACCGTTTCACTCTGATCAATCATTTAGTTACACTAAACCTTGTCACA 720
Db |||||
QY 712 TTGAATGATGGAATATACCGTTTCACTCTGATCAATCATTTAGTTACACCAAACTTGTTCA 771
Db |||||
QY 721 TCTAATGGAATTCAGATTAATAATCAAAATGTCCTGCTGTTATGCTCCATTTATTTGAT 780
Db |||||
QY 772 TCTAATGGAATCTTTATACATATAAAATGTTCTCCCGGTTATGCTCCATTTGTTGAC 831
QY 781 GCTTATATTTCTGCTACAGATGTTAAACCAATATACATTTTAGCATATACCAATGATTATCT 840
Db |||||
QY 832 GCTTATATTTCTGCTACAGATGTTAAATCGTACACCTTCTGCTGATGCTAATGAATATCT 891
Db |||||
QY 841 TGTGCTGGCAGTCTGCTGCAAGTAAACCTTTTCACTTTAAGATGGAATGATCAAGAT 900
Db |||||
QY 892 TGTGCTGGTGGTATTTGGCAACGTCACCTTTTCACTTTAAGATGGAATGATCAAGAT 951
QY 901 AGTAGTGCCGGATCTAACGGTATGTCATTGTTGCTACAACTAGAACAGTTACAGACGT 960
Db |||||
QY 952 AGTAGTGCTGGATCTAACGGTATGTTATTTGGCTACTACAGAACAGTTACAGACGT 1011
QY 961 ACCACTGCTGCTACTACTTTTACCATTCAATCCAAGTGGTTGATAAAACCAAAACAATCGAA 1020
Db |||||
QY 1012 ACTACCGCTGTGACCACTTTACCATTCGATCTTAACCGGACAAACCTAAGACAATTTGAA 1071
QY 1021 ATTTGCAACCTATTCCACCACTACCATCAACCTTATCATATGTTGGTGTGACTTCTCC 1080
Db |||||
QY 1072 ATTTTGAACCTATTCCACCACTACCAATCAACATCATATATGTTGGTGTGACTTCTCC 1131
QY 1081 TATCTGACTTAAGACTGCACCAATGCTGGAACAGCTTACTGTTATGTTGATGTCATAT 1140
Db |||||
QY 1132 TACCTGACCAAACTGCACCAATTTGGGAACTGCTACTGTTATGTTGATATTCATAT 1191
QY 1141 CATACTACCAACTGTTTACCAGTGAATGGACAGGAACAATCACTACCAACCACTCGT 1200
Db |||||
QY 1192 CACACTACCACTACTGTTACCAGTGAATGGACAGGAACAATTACTTCCACCAACAACAT 1251
QY 1201 ACCAATCCAAGTGAATTCATTCACAGTGGTGGTACAAAGTTCCA 1245
Db |||||
QY 1252 ACTAATCCAAGTGAATTCATTCACAGTGGTGGTACAAAGTTCCA 1296

RESULT 7
AF051313
LOCUS
DEFINITION
4383 bp DNA linear PLN 01-AUG-2000
Candida albicans agglutinin-like cell surface protein (ALS8) gene,
complete cds.
ACCESSION
AF051313
VERSION
AF051313.2 GI:9625353
KEYWORDS
SOURCE
ORGANISM
Candida albicans
Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
AUTHORS
Leng, P., Lee, P.R., Wishart, J.A., Wu, H. and Brown, A.J.P.
TITLE
Sequence of the hypha-specific, agglutinin-like cell surface
protein, ALS8 from Candida albicans
JOURNAL
Unpublished
REFERENCE
AUTHORS
Leng, P., Lee, P.R., Wishart, J.A., Wu, H. and Brown, A.J.P.
TITLE
Direct Submission
JOURNAL
Submitted (29-JUL-1999) Molecular and Cell Biology, University of
Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen
AB25 2ZD, UK

COMMENT On Aug 1, 2000 this sequence version replaced gi:4105950.
FEATURES
Location/Qualifiers
1..4383
/organism="Candida albicans"
/mol_type="genomic DNA"
/strain="ATCC10261"
/db_xref="ATCC:10261"
/db_xref="taxon:5476"
566..>3840
/gene="ALS8"
566..>3840
/gene="ALS8"
/product="agglutinin-like cell surface protein"
/note="minor form"
574..579
/gene="ALS8"
/note="E-box"
/bound_moiety="Efg1p"
599..603
/gene="ALS8"
612..>3840
/gene="ALS8"
/product="agglutinin-like cell surface protein"
/note="minor form"
620..>3840
/gene="ALS8"
/product="agglutinin-like cell surface protein"
/note="major form"
630..>3840
/gene="ALS8"
/product="agglutinin-like cell surface protein"
/note="minor form"
697..3840
/gene="ALS8"
/function="involved in cell-cell adhesion"
/note="hypha-specific cell surface glycoprotein; similar
to Candida albicans strain 1161 agglutinin-like protein
encoded by ALS3 gene sequence presented in GenBank
Accession Number U87956; contains two fewer internal
repeated units than the Candida albicans strain 1161 ALS3
gene"
/codon_start=1
/transl_table=12
/product="agglutinin-like cell surface protein"
/protein_id="AAD02580.1"
/db_xref="GI:4105951"
/translation="MLQYTLILLIYLSVATAKTIITGVNFSNLSLTWSNAATHYKGP
TPTWNAVILGSLDGTSPGDTFTLNMPVCFKFTTSQTSVDLTAHGKATCOFQGE
BFMTFTSLTCTVNTLTPSIKALGTVTLLPAPNVGGTSSVDLESKFTAGTNTVF
NDGKKISINVDPFERSNDPKGLTDSRVIPSLNKVSTLFLVAPQCAANGYTSMTGFAN
TYGVDQIDCSNIHVIGITKGLDNWNPVSSPSFYTKTCSNGIFITYKNVPAGYRPFV
DAYISADVNSYTLISANEYTCAGYQWAPPTLWRTGYRNSDAGSNGIVIVATRTV
TDSITATLTFPDRNKTKEILKLPITPTITTSYGVGTSYTKTAPIGETATVI
VDIPHTTITVTSKWTGTTTTHNTPTSDIDTVIVQSPNPVTITTEYWSQSFAT
TTTITPGPNTDVLIREPPNHTVTTEYWSQSFATTTFTTAPPGQTSVITIKBPNP
TVTTEYWSQSFATTTTEYWSQSFATTTTIAAPGDTDTVIIEPPNHTVTTEY
GGDTVLAREPPNHTVTTEYWSQSFATTTTIAAPGDTDTVIIEPPNHTVTTEY
SQSVATTTITAPPGTDLIREPPNHTVTTEYWSQSFATTTTIAAPGDTDTVI
REPPNHTVTTEYWSQSFATTTTIAAPGDTDLIREPPNHTVTTEYWSQSFAT
TVTAPPGDTDTVIIDTWSSSEISFSRPHYNTHTLWSLTWISTKITLETSCSGDK
GCSWVSSTRTVTPNNTPTMTVNTDSTSTSTSTSTSTSTSTSTSTSTSTSTST
TAQTPSVPTTESVEFTTKGNNGPYSPSTHYKSNMDENSEFTTSTAATSDIE
NIATATGVSVEASPPIISSADETTTITTAESTSVIEQPTNNNGGKAPSTSPST
TITANDSVITGTTSTNQSQSQSQSQSDTQTTLTSSQSSLSVLSHMLTTPDGSSSVI
QHSITWICGLITLILSLFI"
BASE COUNT 1359 a 922 c 705 g 1397 t
ORIGIN
Query Match 76.2%; Score 949; DB 8; Length 4383;
Best Local Similarity 85.1%; Pred No. 1e-164;
Matches 1060; Conservative 0; Mismatches 185; Indels 0; Gaps 0;
QY 1 AAGACAATCACTGGTGTGTTTGTATAGTTTAAATCAATCAATGGTCCAATGCTGCTAAT 60

Db 748 AAGCAATCAGTGGTGTTCACACAGTTTAAATTCATTGACCTGGTCTAAATGCTGCTACT 807
Qy 61 TATGCTTTCAAGAGGCGAGGATACCAACTCGGAATGCTGTTTGGGTGGTCTTAGAT 120
Db 808 TATCATATAGGAGCAGGACCCCACTTGAATGCTGTTTGGGTGGTCTTAGAT 867
Qy 121 GGTACAGTGCCAACTCCAGGGGATACATTCATCAATGAATATGCCATGCTGTTAAATAT 180
Db 868 GGTACTAGTCAAGTCCGGGAGATACATTCATCAATGAATATGCCATGCTGTTAAATTT 927
Qy 181 ACTACTTCACAAACATCTGTGATTTAACTGCGGATGGTGTAAATATGCTACTTGTCAA 240
Db 928 ACTACTTCTCAACACATCTGTGATTTGACTGCTCATGGTGTAAATATGCTACTGTCAA 987
Qy 241 TTTTATTTCTGGTGAAGAATTCACAACTTTTCTACATTAACATGCTACTGTGAACGAGCT 300
Db 988 TTTTCAGGAGGTGAAGATTTATGACCTTTCTTCTACATTAACATGCTACTGTGAGCAATCT 1047
Qy 301 TTGAATCATCCATTAAGSCATTTGGTACAGTTACTTTTACCAATTCGCAATCAATGTTGGT 360
Db 1048 TTGACTCCATCTATTAAGGCTTTGGGTACTGTGCACCTTACCACCTTGCATTCATGAGT 1107
Qy 361 GGAACAGGTTCACTCACTGATTTGAAGATTTCTAAATGTTTACTGCTGGTACCAATACA 420
Db 1108 GGAACGTTGTTCTTCTGTTGATTTGAAGATTTCTAAATGTTTACTGCTGGTACTAACACA 1167
Qy 421 GTCACATTTAATGATGGTGATAAAGATATCTCAATTTGATGTTGAGTTTGAAGAAGTCAACC 480
Db 1168 GTTACATTTAATGATGGTGGAAGAAATCTCTAATTTGATGATTTTGAAGAGTCAAT 1227
Qy 481 GTTGATCCAAGTGCAATTTGATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTACA 540
Db 1228 GTCGATCCAAAGGCTACTTAACTGATTCACAGTTATACCAAGTCTCAACAAAGTCA 1287
Qy 541 ACTCTTTTGTGGCACCACCAATCTGCAAAATGTTACACATCTGGTACAAATGGGGTCTCC 600
Db 1288 ACTCTTTTGTGGCACCACCAATCTGCAAAATGTTACACATCTGGTACAAATGGGGTCTCGT 1347
Qy 601 AGTAGTACGGTGACGTTGCTATTCATGCTCAAAATATTCATATTTGTTGATTCACAAAGGA 660
Db 1348 AACCTTATGGTGATGTTCAAAATGACGTTTCAAAATATTCATGTTGTTATTCACAAAGGA 1407
Qy 661 TTAATGATTTGAATTTATCCGGTTTCATCTGAATCATTTAGTTACACTAAACCTTGATCA 720
Db 1408 TTGAATGATTTGAATTTATCCGGTTTCATCTGAATCATTTAGTTACACTAAACCTTGATCA 1467
Qy 721 TCTAATGGAATTCAGATTAATATCAAAATGATACCTGCTGGTTATCGTCCATTTATGAT 780
Db 1468 TCTAATGATTTGAATTTATCCGGTTTCATCTGAATCATTTAGTTACACTAAACCTTGATCA 1527
Qy 781 GCTTATATTTCTGCTACAGATGTTAAACCAATATATCTTTAGCATATACCAATGATTTACT 840
Db 1528 GCTTATATTTCTGCTACAGATGTTAACTCGTACACTTGTGCTGATGCTAATGAAATATCT 1587
Qy 841 TGTGCTGGCAGTCTGCTGCAAAAGTAAACCTTTTCACTTTAAGATGCACTGGATCAAGAAT 900
Db 1588 TGTGCTGGGTTATTTGGCAAGCTGCACCTTTTCAATTAAGATGCACTGGATCAAGAAT 1647
Qy 901 AGTGATGCGGATCTAAACGGTATGTCATTTGTTGCTTACAACTTAGAACAGTTTACAGACAGT 960
Db 1648 AGTGATGCTGGATCTAAACGGTATGTTTATTTGGTGGCTACTTACCAGAACAGTTTACAGACAGT 1707
Qy 961 ACCACTGCTGTCACTACTTTACATTTCAATTCATCAAGTGTGTAATAACCAACCAATCGAA 1020
Db 1708 ACTACCGCGGTGACCACTTACATTCGATCTCTAAACCGGACCAAACTTAAGCAATGAA 1767
Qy 1021 ATTTTGCACCTTATTCACCACTACTACCACTCAAACTTATATATGTTGGTGTGACTACTTCC 1080
Db 1768 ATTTTGAACCTTATTCACCACTACTACCACTCAAACTATACATCATATGTTGGTGTGACTACTTCC 1827
Qy 1081 TATCTGACTAAGACTGCACCAATTTGGTGAACAGCTACTGTTATTTGTTGATGTCGCATAT 1140

Db 1828 TACCTGACCAAACTGCACCAATTTGGGGAACCTGCTACTGTTATTTGTTGATATTCATAT 1887
Qy 1141 CATACTACCAAACTGTTTACCACTGAATGGAGCAACAACTACTACCAACCAACTGCT 1200
Db 1888 CACACTACCACTACTGTTTACCACTGAATTTGGAGCAACAACTACTTCCACCAACACAT 1947
Qy 1201 ACCAATCAACTGATTCAATTTGACACACTGGTGGTACAAGTTCCA 1245
Db 1948 ACTAATCCAACCTGCTCAATAGACACTGCTATTGTACAAGTTCCA 1992
RESULT 8
CAU87956
LOCUS CAU87956 3360 bp DNA linear PLN 02-JUL-1998
DEFINITION Candida albicans agglutinin-like protein (ALS3) gene, complete cds.
ACCESSION U87956
VERSION U87956.1 GI:3273414
KEYWORDS
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 3360)
AUTHORS Hoyer, B.L., Payne, T.L., Bell, M., Myers, A.M. and Scherer, S.
TITLE Candida albicans ALS3 and insights into the nature of the ALS gene family
JOURNAL Curr. Genet. 33 (6), 451-459 (1998)
MEDLINE 98309840
PubMed 9644209
REFERENCE 2 (bases 1 to 3360)
AUTHORS Hoyer, B.L.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-1997) Veterinary Pathobiology, University of Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES
source
1..3360
/organism="Candida albicans"
/mol_type="genomic DNA"
/strain="1161"
/db_xref="taxon:5476"
/chromosome="R"
/map="S+U"
1..3360
/gene="ALS3"
1..3360
/gene="ALS3"
/codon_start=1
/transl_table=12
/product="agglutinin-like protein"
/protein_id="AAC39486.1"
/db_xref="GI:3273415"
TPTWNAVLGWSLDGTSASPGDTFLNCPVFKFTTSQTSVDLTAGVKYATCQFOAGE
EFMTFSLTCTVSNLTLPKALGTVTLPLAFNVGSGTSSVDLESKCTAGTNTVTF
NDGKKLSINVDPERSNVDPKGYLTSRVPISLNKSTLFPVAPQANGYTGTMGFAN
TYGQVQDCDSNIHVGTIKGLDMNYPVSSSFYTKTCSNGIFITYKNVPAGYRPFV
DAYTSATDVNSYTLSEYNEAYTCAGYORAPFTLRWTCYRNSDAGSNGIVATRTV
TDSATVTLTLPDNRDKTKTIEILKPIPTTITTSYVGVTTSTKTAPEGTDVLI
VDIYHTTITVSKWTGITSSTHTHTDSDIDVIVQVSPNPVTTTTEWSQSFAT
TTTTTGPGNTDVLIREPPNHTVTTTEYWSYTTTSTFTTAPPGTDSVLIKEPPNP
TVTTTEWSYTTTSTFTTAPPGTDSVLIKEPPNHTVTTTEYWSQSYTTTFTTAP
GGTDVLIAREPPNHTVTTTEYWSYTTTFTTAPPGTDSVLIAREPPNHTVTTTEY
WSYATTTITAPPGTDVLIAREPPNHTVTTTEYWSYATTTITAPPGTDVLI
REPNNHTVTTTEYWSQSFATTTTAPPGTDVLIAREPPNHTVTTTEYWSQSYATTT
TITAPPGTDVLIAREPPNHTVTTTEYWSQSYATTTITAPPGTDVLIAREPPNHT
TITAPPGTDVLIAREPPNHTVTTTAPPGTDVLIAREPPNHTVTTTAPPGTDVLI
EKTITETSCGDKGCSVSTRIVIPNNITPMVTNTVDSTSTSTSTSTSTSTSTST
ESGVSTVETSTVTAQNPSTVETSTVETSTVETSTVETSTVETSTVETSTVETST
PTTAASTSTDIENETATGSEVSPISISSADETTTVAESTSVTEQNTNN
GGKAPASATSSPTTTANNDVLTGTTSTNQSQSQSQSQSQSQSQSQSQSQSQSQSQ
LHMLTTFDGSQSVIOHSTWLCGLTLTLFLPI"
BASE COUNT 1029 a 804 c 520 g 1007 t
ORIGIN

/db_xref="GI:3598670"
/translation="MLLPLLLSLCSVATAKVIITGVCFNSFDSLTRAGNAYAKGPN
RPTNVALGMSLDGTSPNGDFTFLNMPGVKFIITDQTSVDLTAGSCVACQYVSGE
ETTFSSSLKQTNLITSLIKALGTIVLPIFENVGGTSSVDLESQCFKACTNTVTF
NOGDKLSIDVDFEKTNEBDAQFYIASHLIPSIKVKYTIYVAPQACANGYTSAMGFI
LTGDTTIDCSNHWNPFIKGLNDWNPVSSDSLSYNKTSCTGISITYENVPVAPRPF
DYVTSYGGNQRLTYNDYACVGSLSQSKPFLRLRGYNNSSANGSIYVATTITV


```

TPTWTAIVGWSLDGATASAGDITFTLDMPCVKFIIDOTSIDLVDGRYATCNLSAE
BFTTSSSVCTVTTMTADTKAIGTVTLFPFSGSGSDVDLANSQCFTAGINTVTF
NDGTSISTVDVDFKSTVASSDRILLRILPSLSOAVNLELPOECANGVTSCTWGFST
AGTGATIDCSVTHVQISNGLNDWNPISSEFSYTKTCTSTSLVLTFRONVPAGYRPFV
DAYISATRVSSYTMQIYACVGAASVDDSFTHWRGTSNSQAGSNQIIVVTRTV
TDSSTAVTLTPNSDTRKTEILLOPIPTTTTTSYVGVTSYSTKAPIGETATVI
VDVPHVTTTTSVSEWGTGTTTTTTRTNPDSIDTVVQVPSNPVTVTITTEWSQYAT
TTTVTAPPGGDSVIREP"
repeat_region 1300..1407
/note="first of multiple copies of 108bp imperfect repeat"
BASE COUNT 391 a 294 c 253 g 469 t
ORIGIN
Query Match 58.9%; Score 733.4; DB 8; Length 1407;
Best Local Similarity 74.5%; Pred. No. 5.5e-125;
Matches 923; Conservative 0; Mismatches 316; Indels 0; Gaps 0;
QY 7 ATCACTGGTGTGTTTTCATAGTTTAAATTCATTAACATTTGGTCCAATGCTGCTAAATTAATGCT 66
DB 58 ATTACAGGTGTTTTCATAGTTTAAATTCGTTAACTTTGGCCCAATGCTGCTTCTTATCCA 117
QY 67 TTCAAAGGCCAGGATACCCAACTTGGAAATGCTGTTTGGGTTGGTCTTCTAGATGGTACC 126
DB 118 TATAGAGTCCAGTACTCTACTTGGACCGCTGTAATAGGATGGTCTTTAGATGGAGCT 177
QY 127 AGTGCCAATCCAGGGATACATTCATTAATGATATGCCATGCTGTTTAAATATACTACT 186
DB 178 ACTGCTAGTGGTGGTGACATTCACGTTAGACATGCGCTTGTGTTTTCAAATTTTATCT 237
QY 187 TCACAAACATCTGTTGATTTAACTGCCGATGGTGTGTTTAAATATGCTACTTGTCAAATTTAT 246
DB 238 GATCAAAACGTCATTTGATTTAGTTGCTGATGGTCTGCTACTTATGCTACTTGTAAATTTGAAT 297
QY 247 TCTGGTGAAGATTCACAACTTTTCTACATTAACATGACTGTGTGAACGACGCTTTGAAA 306
DB 298 TCTGCCGAAGATTTACTACTTTTCTAGTGTGTCATGACTGTGACTACTACTACTCAATGACT 357
QY 307 TCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTTGCAATTCATGTTGGTGGMACA 366
DB 358 GCTGACACCAAGGCCATAGGAATCTGAACATTAACCTTTCTCAATTCAGTGTGGGGGATCA 417
QY 367 GGTTCATCAACTGATTTGAAGATCTCAATGTTTACTGCTGGTACCAATACAGTCACA 426
DB 418 GGTTCAGATGTTGATTTGCAAAATCTCAATGTTTACTGCGAGAAATCAATACAGTTACT 477
QY 427 TTTAATGATGGTGATAAAGATATCTCAATTTGATGTTGAGTTTGAAGAAGTCAACCGTTGAT 486
DB 478 TTTAATGATGGTGACACTAGCATTTCCACACAGTTTGAATTTTGAAAAATCAACCGTGGCC 537
QY 487 CCAAGTGCAATTTGATGCTTCCAGAGTTATGCCAAGTCTCAATTAAGTCAACTCTTT 546
DB 538 TCACGGGATCGTATCTTGTGTGCAAGAAATTTTACCAGTCTTTCAACAGCAGTAAATCTT 597
QY 547 TTTGTGCGACCAATGTCGAATGTTTACACATCTGTTACAAATGGGTTTCCAGTAGT 606
DB 598 TTTCTTCCCAAGAAATGTCGAATGTTTATCTTCTGTTACAAATGGGATTTTCCAGTGT 657
QY 607 AACGGTGACGTTGCTATTTGCTCAAAATTTATCATTTGGTATCAAAAAGGATTAAT 666
DB 658 GGTACTGGTGCTACTATAGATGTTCCACAGTTTATCGGATATCAAAATGGTGAAT 717
QY 667 GATTGAATATCCGGTTTCACTGAAATCATTTAGTTTACATTAACCTTTGATCAATTAAT 726
DB 718 GATTGGAATATTCAAATTTTCACTGGAATCTTTTCTTACAAAAGACCTGTACATCAACA 777
QY 727 GGAATTTACAGATTAATAATCAAAATGACTGCTGTTTATGCTCCATTTATGATGCTTAT 786
DB 778 AGTGTTTATGATTAATTTTCAAAATGTTCTGCGGATATGCTCCATTTGATGCTTAT 837
QY 787 ATTTCTGCTACAGATGTTTAAACCAATATCTTTTAGCATATACCAATGATTAATCTTGCT 846
DB 838 ATTTCTGCAACAGGAGTCACTCATATACCATGCAATACATTAATATATATATGCTTGT 897
```

```

847 GCAGTCTGTCTGCAAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGAT 906
DB 898 GCGCGGCTTCTGTGATGACTCATTTACTCATCTTGGCGGGATATAGTAATAGTCAA 957
QY 907 GCGGATCTACGGTATTTGTCATTTGCTTCAACTAGAACAGTTTACAGACAGTACCCT 966
DB 958 GCTGGTCTTAATGGTATTTACCAATTTGGTAAACAACATAGAACAGTTTACAGACAGTACCCT 1017
QY 967 GCTGTCATCTTACCAATTTCAATTTCAATTTGATTAATAACCAAAACCAATCGAAATTTTG 1026
DB 1018 GCTGTGACTTCTTACCAATTTCAATTTCCGATCTGACAAACCAAAACCAATCGAAATTTTA 1077
QY 1027 CAACCTATTTCCAAACCACTACCATCACAACTTTCATATTTGTTGGTGTGACTACTTCTTCTATCTG 1086
DB 1078 CAACCTATTTCCAAACCACTACCATCACAACTTTCATATTTGTTGGTGTGACAACTTCTTCTACCTG 1137
QY 1087 ACTAAGACTGCACCAATTTGGTGAACAGCTACTGTTATTTGTTGATGTGCCATATCATACT 1146
DB 1138 ACTAAAACCTGCACCAATTTGGTGAACAGCTACTGTTATTTGTTGATGTGCCATATCATACT 1197
QY 1147 ACCCAACTGTTTACCAATTTGAGTGGACAGGAACAATCACTACCAACCACTCGTACCCT 1206
DB 1198 ACTCAACTGTTTACCAATTTGAGTGGACAGGAACAATCACTACCAACCACTCGTACCCT 1257
QY 1207 CCAACTGATTTCAATTTGACACAGTGGTGGTCAAAAGTTCCA 1245
DB 1258 CCAACTGATTTCTATAGACTGTCGTTGTTCAAGTTCCA 1296

RESULT 13
AF272027 4569 bp mRNA linear PLN 24-OCT-2000
LOCUS Candida albicans agglutinin-like protein mRNA, partial cds.
DEFINITION AF272027
ACCESSION AF272027 GI:10952735
VERSION AF272027.1
KEYWORDS
SOURCE
ORGANISM
Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE
1 (bases 1 to 4569)
AUTHORS Chen,X. and Chen,J.-Y.
TITLE ALS4 (agglutinin-like sequence) of Candida albicans
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 4569)
AUTHORS Chen,X. and Chen,J.-Y.
TITLE Direct Submission
JOURNAL Submitted (24-MAY-2000) Shanghai Institute of Biochemistry, Yue
Yang Road 320, Shanghai 200031, China
FEATURES
source
1..4569
/organism="Candida albicans"
/mol_type="mRNA"
/db_xref="SC5314"
/db_xref="taxon:5476"
1..>4569
/codon_start=1
/transl_table=12
/product="agglutinin-like protein"
/protein_id="AAG25054.1"
/db_xref="GI:10952736"
/translation="MLQLFLLSLCVSVATAKVIIGIPDSFNSLTWTNAASYSYRPGA
NPTWTAIVGWSLDGATASAGDITFTLDMPCVKFIIDOTSIDLVDGRYATCNLSAE
BFTTSSSVCTVTTMTADTKAIGTVTLFPFSGSGSDVDLANSQCFTAGINTVTF
NDGTSISTVDVDFKSTVASSDRILLRILPSLSOAVNLELPOECANGVTSCTWGFST
AGTGATIDCSVTHVQISNGLNDWNPISSEFSYTKTCTSTSLVLTFRONVPAGYRPFV
DAYISATRVSSYTMQIYACVGAASVDDSFTHWRGTSNSQAGSNQIIVVTRTV
TDSSTAVTLTPNSDTRKTEILLOPIPTTTTTSYVGVTSYSTKAPIGETATVI
VDVPHVTTTTSVSEWGTGTTTTTTRTNPDSIDTVVQVPSNPVTVTITTEWSQYAT
TTTVTAPPGGDSVIREP"
CDS
```

Qy	1	AAGACAATCACTGGTGTGTTTTTGATAGATTTTAAATTCATTAACTGGTCCAAATGCTGCTAAT	60
Db	52	AAGGTTATTACGGGCAATTTTTTCGATAGATTTTAAATTCGTTAACTTGGACCAATGCTGCTCTCT	111
Qy	61	TATGCTTTCAAGGGCCAGGATACCCAACTTGGNAATGCTGTTTTGGGTTGGTCCCTAGAT	120
Db	112	TATTCATATAGAGGTCCAGCTAAATCCCTACTTGGACCGCTGTAATAGAGATGGTCTTTAGAT	171
Qy	121	GGTACCAGTGCMAATCCAGGGGATACATTCACATTTGAATATGCCATCTGTGTTTTAAATAT	180
Db	172	GGAGCTACTGCTAGTGTGGTGACACAATTCACGTTAGACATGCCAATGTGTTTCCAAATTT	231
Qy	181	ACTACTTCCAAAAATCTGTGTGATTTTAACTGCCGATGGTGTAAATATGCTACTTGTCAA	240
Db	232	ATTACTGATCAAAACGTCAAATTTAGTTAGTTGCTGATGTCGTAATGCTACTTTGTAAT	291
Qy	241	TTTTTATTTCTGGTGAAGAATTCACAACTTTTTTCTACATTAACATGTACTGTGAACGACCT	300
Db	292	TTGAACCTCTGCGGAAGAGTTTACTACTTTTTCTAGTGTGTCATGTACTGTGCACTACTACA	351
Qy	301	TTGAAATCATCCATTAAGGCATTTGGTACAGTTACTTTTACCAATTCGATTCAAATGTTGGT	360
Db	352	ATGACTCTGACACCAAGCCATAGGAACGTAAACATTACTCTTTCTCATTCAGTGGGG	411
Qy	361	GGAAACGGTTTCATCAACTGATTTTGGAAAGATTCCTAAATGTTTTACTGCTGGTACCATAACA	420
Db	412	GGATCAGGTTCCAGATGTTGATTTTGGCAAAATTCCTCAATGTTTTCTGCAGGAATCAATACA	471
Qy	421	GTCACATTTAATGATGGTGTAAAGATATCTCAATTTGATTTGAGTTTGGAAAGTCAACC	480
Db	472	GTTACTTTTAAATGATGGTGACACTAGCAATTCGCCAACAGTTGATTTTGGAAATCAACC	531
Qy	481	GTTGATCCAAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAGTCTCAATAAAGGTCAACA	540
Db	532	GTGSCCTCCAGCGATCGTATCTTGTGTTCAAGAAATTTTACCCAGTCTTTTACAAAGCAGTA	591
Qy	541	ACTCTTTTGTGGCACCAATGTGAAATGGTTTACACATCTGGTACAAATGGGGTTCTCC	600
Db	592	AGTCTTTTCTTCCCAAGAAGTGTCAAAATGGTTTATACCTTCTGGTCAAAATGGGAATTTTCA	651
Qy	601	AGTAGTAACCGTGAGTTGCTATTGATTTGCTCAAAATATTCATATTTGGTATGCACAAAAAGGA	660
Db	652	ACTGCTGGTACTGGTGTACTATATAGATTTGTTTCCAGTTTATGTCGGGATATCAAAATGGG	711
Qy	661	TTAAATGATTTGAATTAATCCGGTTTTCATCTGAATCATTTAGTTTACACTTAAACCTTGTACA	720
Db	712	TTGAATGATTTGAATTAATCCAAATTTCTACTGGAATCTTTTCTTACACAAGACCTGTACA	771
Qy	721	TCTAATGGAATTCAGATTTAAATATCAAAATGTACCTGCTGGTTATCTGTCATTTATTGAT	780

BASE COUNT	292 a	205 c	176 g	334 t	
ORIGIN					
					protein_id="AAG35623.2"
					/db_xref="GI:11693480"
					/translation="ASPGDTFTLLMPCVKFIFTQTSVDLTANGVKYATCSFYSGBEF
					TTFSSTCTVNSALTSVSRAPGTVTLPISPNVCGTGSVDLSDSKCFTAGTWTVTID
					GDNKSTVDFDEKSTVDGYLTSRLMPSLNKVTSLFVAPQCARGYTSGTGIFSSN
					GGVFSFCNSVHVGTGVNDNFWPVSSEFSYTKTCSGTITTYRNPAGYRPFDA
					YITASDVNSYTLGYNDYTCVGGSVQHKPFLRWGSKNGEAGSGNIVATVTRTVTD
					STAVTITLPNSVDVKTIETIIQPIPTTITLYGVGTTSISYTKTAPIGDATTATLFD
					IP"

Query Match	58.8%;	Score 731.8;	DB 8;	Length 1007;
Best Local Similarity	82.9%;	Pred. No. 1.2e-124;		
Matches	835;	Conservative 0;	Mismatches 172;	Indels 0; Gaps 0;
Qy	130	GCCAACTCCAGGGATACATTCACATTCAGAAATATGCGCATGTGTTTAAATATACACTACTTCA	189	
Db	1	GCAAGTCCGGAGACACATTTACATGATCATGCTTGTTGTTTCAAAATTAACACTTACC	60	
Qy	190	CAAAACATCTGTTGATTTTAACTGCCGATGGTGTAAATATGCTACTTGTCAATTTTATTTCT	249	
Db	61	CAAAACATCTGTTGATTTTGA CGGCAATGGTGTAAATATGCTACATGCTGTTTATTTCG	120	
Qy	250	GGTGAAGAAATTCACAACTTTTCTACATTAACATGTA CTGTGAACGACGCTTTGAAATCA	309	
Db	121	GGTGAAGAAATTCACAACTTTTTCATCATTAACATGTACTGTAAACAGTGTCTTTAAACCTCA	180	
Qy	310	TCCATTAAGGCATTTGGTACAGTTACTTTTACCATTTGCCATTCATTCATGTGTGGTGAACAGGT	369	
Db	181	TCGTGTTAAGGCTTTTGGTAGCCGTTACTTTTACCATCTCATTCATGTAGTGGTGAACGGGA	240	
Qy	370	TCATCAACTGATTTTGAAGAGATTCFAAATGTTTACTGCTGCTGATCCAATACAGTCACATTT	429	
Db	241	TCGTGAGTGTGATTTAGAAGACTCCAAATGCTTCACTGCTGATCCAACCGTCATCTTT	300	
Qy	430	AATGATGGTGATAAAGATATCTCAATTTGATGTGAGTTTGAAAGTCAACCGTTGATGCCA	489	
Db	301	ACCGACGGGACACAAAGTCTCAACTACTGTTGATTTTGAAAAATCAACAGTTGATTTCA	360	
Qy	490	AGTGCATATTTGTATGCTTCCAGAGTTTATGCCAAGTCTCAATAAGTGCACAACTCTTTTT	549	
Db	361	ACTGGATATTTGACTTCCCTCCAGACTTTATGCCAAGTCTCAATAAAGTCACATCTCTTTTT	420	
Qy	550	GTGCGACACCAATGCTGAAAAATGGTTCACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC	609	
Db	421	GTACACCAACAAATGTCAGAGGTTTATACATCTGGTACAATGGATTTTTCAGTAGTAGTAA	480	
Qy	610	GGTGACGTTGCTATTGATTTGCTCAAAATTTATCATATTGATATCAACAAAGGATTAATATGAT	669	
Db	481	GGTGGTGTCTCTTTTGACTGTTTCAAAATGTTTCATGTTGGAATTAACAAATGGTGTGAATGAT	540	
Qy	670	TGGAATTAATCCGGTTTCATCTGAATTCATTTAGTTTACACTTAAACTGTGTACATCTAAATGGA	729	
Db	541	TGGAATTTTCCAGTGTCTACAGAAATCAATTTAGTTTACACTTAAACTGTGTCTACAGGT	600	
Qy	730	ATTCCAGATTAATATCAAAATGATACCTGCTGTTTATCGTCCATTTTATTTGATGCTTATATT	789	
Db	601	ATTAACAATTCATACAGAAATGTTTCCCTGCTGGTTATCGTCCATTTATTCGACGCTTACATT	660	
Qy	790	TCTGCTACAGATGTTTAAACCAATATACATTTTAGCATATACCAATGATTATATCTGTGCTGGC	849	
Db	661	ACTGCTTCGGATGTTTAACTGTATACCTTGTGCATATACCAATGATTATCTGTGTTGGT	720	
Qy	850	AGTCGTCTGCAAAAGTAAACCTTTTCACTTTTAAGATGGACTGGATACAAAGATATAGTGATGCC	909	
Db	721	GGTAGCGTGCACATAAACCATTTTACTTTTAAGATGGTCTGGATACAAAAATGGGGAAGCC	780	
Qy	910	GGATCTAAACGCTATTTGTTCATTTGTTGCTACAACTAGAACAGTTTACAGACAGTACCACGTCT	969	
Db	781	GGGTCTAAACGCTATTTGTTCATTTGTTGCTACAAACAAAGAACAGTTTACAGATAGTACTGTCT	840	
Qy	970	GTCACTACTTTTACCATTCAATCCAAAGTGTTCGATTAACCAAAAAACAAATTCGAAATTTTGCAA	1029	

Matches 920; Conservative 0; Mismatches 319; Indels 0; Gaps 0;			
Qy	7	ATCACTGGTGTGTTTTCATAGTTTAAATTCATTAACTTGGTCAATGCTGCTAAATATGCT	66
Db	58	ATTACAGGTGTTTTCATAGTTTAAATTCATTAACTTGGGCCAATGCTCTTCTTATCCA	117
Qy	67	TTCAAAGGCGCAGGATPACCAACTTGGGAATGCTGTTTGGGTGGTCTTAGAGTGAC	126
Db	118	TATAGAGGTCACGCTACTCTCTTGGACCGGTGTAATAGGATGGTCTTTAGATGGAGCT	177
Qy	127	AGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATATCTACT	186
Db	178	ACTGCTAGTGTGCTGGGACACATTCACGTTAGACATGCCCTTGTTGTTTCAAAATTTATCT	237
Qy	187	TCACAAACATCTGTTGAATTAATGCGGATGGTGTAAATATGCTACTTGTCAAAATTTAT	246
Db	238	GATCAACGCTCAATTCGATTTAGTTGCTGATGGTCTGCTACTTATGCTACTTGTAAATTTGAAC	297
Qy	247	TCGTGTAAGAATTCACAACTTTTCTACATTAACATGCTACTGTGTAACGACGGTTTGA	306
Db	298	TCGCGCAAGAGTTTACTACTTTTCTAGTGTGTCATGCTACTGTGCTACTACTACAATGACT	357
Qy	307	TCATCCATTAAGCATTTGGTACAGTTACTTTACCAATTTGCATTCAATGTTGGTGGAA	366
Db	358	GCTGACACCAAGCCCATGGAACTGTAACTATACCTTTCTCAATTCAGTGTGGGGGATCA	417
Qy	367	GGTTTCATCACTGATTTGGAAGATCTAAATGTTTACTGCTGGTACCAATACAGTCACA	426
Db	418	GGTTCAGATGTTGATTTGGCAATTTCTCAATGTTTACTGCGAGGAATCAATACAGTTACT	477
Qy	427	TTTAATGATGGTGATAAAGATATCTCAATTTGATGTTGAGTTTGAAGAAGTCAACCGTTGAT	486
Db	478	TTTAATGATGGTGACACTAGCATTTCCACACAGTTTGAATTTTGAAGAAATCAACCGTGGCC	537
Qy	487	CCAAGTGCATATTTGATGCTTCCAGAGTTATGCCAAGTCTCAATTAAGTCAACTCTT	546
Db	538	TCCAGCGATCGTATCTTGTGTTGTCAGAAATTTTACCAGTCTTTTCAAGCAGTAAGTCTT	597
Qy	547	TTTGTGGCACCAATGTGAAATGTTACATCTGTGTACATCTGGGTCTTCCAGTAGT	606
Db	598	TTTCTTCCCAAGATGTGCAATGGTTATATCTTCTGGTACATGGGATTTTCAACTGCT	657
Qy	607	AACGGTGACGTTGCTATGATGCTCAAAATTTCAATATTTGATGTTATCAAAAGGATTAAT	666
Db	658	GGTACTGGTGTACTATAGATTTGTTCCACAGTTCATGTCGGGATATCAAAATGGGTTGAAT	717
Qy	667	GATTGGAAATATCCGGTTTCTCTGNAATCATTTAGTTACACTTAAACCTTGTACATCTAAT	726
Db	718	GATTGGAATTTATCCAAATTTTCACTGGAATCTTTTCTTACAAAGACCTGTACATCAACA	777
Qy	727	GGAAATTCAGATTAATATCAAAATGTACCTGTCTGTTATCGTCCATTTATTCATGCTTAT	786
Db	778	AGTGTTTTAGTAACCTTATCAGAAATGTTCTGCGGATATCGTCCATTTGTCGATGCTTAT	837
Qy	787	ATTTCTGCTACAGATGTTAAACCAATATATCTTTAGCATATPACCAATGATTTATCTTGTCT	846
Db	838	GTTTCTGTCAACACGAGTACGCTCATATACCATGCAATACACTAATATATATGCTTGTGTT	897
Qy	847	GGCAGTCTGTCGAAGTAAACCTTTTCACTTTAAGATGGACTGGATACAAGAATAGTGAT	906
Db	898	GGCGGGGCTTCTGTTGATGACTCATTTACTACTACTTGGCGGGGATATAGTAATAGTCAA	957
Qy	907	GGCGGATCTAACGGTATTTGCTGCTATGCTGCTGCTGTTATCGTCCATTTATTCATGCTTAT	966
Db	958	GCTGGTCTTAATGGTATTAACCAATTTGGTAACTAGAACAGTTTACAGACAGTACCCT	1017
Qy	967	GCTGTCACTACTTTTACCAATTCAGTGTGTGATTAACCAAAACCAATCGAAATTTTG	1026
Db	1018	GCTGTGACTACTTTTACCAATTCAGTGTGATTAACCAAAACCAATCGAAATTTTA	1077
Qy	1027	CAACCTATTTCCAAACCACTACCATCAACTTCAATGTTGGTGTGACTACTTCTTATCTG	1086
Db	1078	CAACCTATTTCCAAACCACTACCATCAACTTCAATGTTGGTGTGACTACTTCTTATCTG	1137

Search completed: January 17, 2004, 22:42:32
Job time : 4790 secs

Qy	1087	ACTAAGACTGCACCAATTTGGTGAACAGCTACTGTTATTTGTTGATGTTGCCATATCATACT	1146
Db	1138	ACTAAACTGCACCAATTTGGTGAACAGCTACTGTTATTTGTTGATGTTGCCATATCATACT	1197
Qy	1147	ACCACAACTGTTTACCAGTGAATGGACAGGAACAATCATTACCAACCAACTCGTACCAAT	1206
Db	1198	ACTACAACTGTTTACCAGTGAATGGACAGGAACAATTTACTTACCACTACCACTCGTACCAAT	1257
Qy	1207	CCAACTGATTTCAATTTGACACAGTGGTGTACAAAGTTCCA	1245
Db	1258	CCAACTGATTTCTATAGATACTGCTGTTGTTCAAGTTCCA	1296

